Yemima Ben-Menahem

CAUSATION IN SCIENCE

CAUSATION IN SCIENCE

Causation in Science

YEMIMA BEN-MENAHEM

PRINCETON UNIVERSITY PRESS
PRINCETON & OXFORD

Published by Princeton University Press,
41 William Street, Princeton, New Jersey 08540

In the United Kingdom: Princeton University Press,
6 Oxford Street, Woodstock, Oxfordshire OX20 1TR

press.princeton.edu

ISBN 978-0-691-17493-8
Library of Congress Control Number: 2018932375

British Library Cataloging-in-Publication Data is available

This book has been composed in Arno Pro

Printed on acid-free paper. ∞

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

In memory of my parents
Elizabeth and Joseph Goldschmidt

CONTENTS

PREFACE

THE CONCEPTS OF TRUTH AND CAUSATION, generally taken for granted in daily discourse, have engaged philosophers throughout history. With regard to both these concepts, the disparity between appearance and reality, and the difficulty of clearly delineating one from the other, threaten to subvert the uncritical stance of common sense. Moreover, in discourse about causation, as in discourse about truth, human language, with its categories, descriptions, ambiguities, and metaphors, seems to stand in the way of an objective, mind-independent grasp of reality. This obstacle, though recognized by earlier philosophers, came to the fore at the turn of the twentieth century, when developments in logic, mathematics, science, and philosophy converged into the philosophical orientation known as the "linguistic turn." Radical manifestations of the concern about language's role in epistemology went so far as to call for the rejection of truth and objectivity, and their replacement by notions such as definitions, conventions, fictions, and narratives. This linguistic challenge to truth (the subject of my *Conventionalism*) will not concern us here.

The linguistic turn is also evident in twentieth-century philosophy's focus on the theory of meaning in general, and on the meanings of certain core concepts, such as causation, explanation, virtue, and liberty. The importance of these meaning-oriented projects notwithstanding, in the case of causation, the ongoing debates about its definition have diverted attention from other issues, and in particular, from the various ways in which causal notions function in contemporary science. Attention to scientific practice led me to the conclusion that the notion of causal constraint is far more germane than that of causal relations between individual events. This book is therefore structured around a fam-

ily of general constraints encountered in fundamental science: determinism, locality, stability, symmetries, conservation laws, and variation principles. Its chapters examine the interrelations between members of this family of constraints—between locality and determinism, determinism and stability, and so on. My treatment of these subjects does not aspire to completeness; some constraints, such as the asymmetry of the causal relation, are deliberately omitted, and in any case, the list of causal constraints is as open-ended as contemporary science in general. As conceived here, causal constraints do not have an a priori basis; even when they arise from deeply rooted intuitions, they are part of science, that is, they must have empirical support, and are always subject to reevaluation.

I understand causal constraints as general constraints on change. As such, they constitute the conceptual scaffolding of the natural sciences, and differ from purely mathematical constraints, which are indifferent to temporal change and temporal evolution. The causal-constraint approach to causation has a significant advantage over the traditional approach. A key goal of the scientific enterprise is to explain not only that which occurs, but also that which is excluded from occurring. While both occurrences and exclusions can be explained by causal constraints, the traditional approach mainly focuses on the former, rendering its explanatory capacity limited. Hence, the broader understanding of causation proposed here.

Yet this broader conception does not lessen the need to take into account the role of language in representing reality. As Davidson demonstrated, even if causal relations between particular events are conceived as independent of the descriptions chosen to describe these events, when such causal relations are adduced in an explanatory theory, they lose this description independence. I take causal discourse, in the traditional sense as well as the broader sense championed here, to be objective. At the same time, where relevant, I seek to be mindful of description sensitivity. This sensitivity has implications both for our understanding of specific theories, for example, statistical mechanics, and for the problem of intertheoretic relations in general, discussed in the book's concluding chapter.

The place of causation in science is controversial. Some philosophers argue that causal discourse should be eliminated, while others find it useful at the fundamental level of physics, but maintain that higher-level theories, being reducible to this fundamental level, make no causal claims of their own. In drawing attention to the spectrum of causal constraints that guide fundamental science, the argument set forth in this book takes issue with causal eliminativism. In elucidating the structure of intertheoretic relations, it challenges causal reductionism.

ACKNOWLEDGMENTS

I WOULD LIKE TO THANK Meir Hemmo, Daniel Rohrlich, Orly Shenker, and the other members of the Foundations of Physics group at the Hebrew University of Jerusalem's Edelstein Center for the History of Science, Technology, and Medicine, for many thought-provoking discussions. Conversations with Haim Sompolinsky while teaching a joint seminar on physics, neuroscience, and free will were similarly stimulating. I am grateful to Peter Dougherty, Al Bertrand, and Leslie Grundfest of Princeton University Press for their support and assistance, to the anonymous readers for their many helpful suggestions, and to Nessa Olshansky-Ashtar for improving the language and style. Springer, Wiley-Blackwell, and *Iyyun* kindly permitted me to reproduce material from previously published work. Specific acknowledgments appear in the relevant places. Itamar Pitowsky and Bill Demopoulos, who are no longer with us, were a source of inspiration and encouragement. Hanina Ben-Menahem's critique, combined with his unfailing trust and support, was vital; I cannot thank him enough. This book is dedicated to the memory of my parents, Elizabeth and Joseph Goldschmidt, who lived through precarious times but were fortunate enough to be able to leave Germany and build a life in Israel, pursuing their vocations in science and education with great devotion and integrity.

CAUSATION IN SCIENCE

1

From Causal Relations to Causal Constraints

THIS BOOK EXAMINES the family of causal notions and causal constraints employed in fundamental science, and analyzes some of the conceptual relations between them. It argues that the concepts of determinism, locality, stability, and symmetry, as well as conservation laws and variation principles, constitute a complex web of constraints that circumscribe the causal structure of our world. It argues, further, that mapping out the various links between these causal constraints is an indispensable, though neglected, aspect of the project of understanding causation. The book thus seeks to shift our attention from causal relations between individual events (or properties of events) to the more general causal constraints found in science, and the relations between them. In so doing, it does not purport to replace causal relations with causal constraints in every context, but rather to suggest a broader perspective on causation and a new research program for the philosophy of causation.[1]

Philosophical analysis of complex concepts usually begins with definitions. The exploration of causality is no exception. Enormous effort

1. I take "causation" and "causality" to be synonymous, generally using the latter when referring to writings that use this term, and the former otherwise. As I explain, I extend the application of these terms beyond a relation between individual events, and hence they cannot always be associated with specific cause-events or effect-events. The same caveat applies to the notion of causal *constraint*, which is the focus of this book.

has been devoted to formulating the "right" definition of causation and defeating rival definitions. Regularity theories, counterfactual analyses, interventionist/manipulation accounts, probabilistic theories, transmission accounts, and explanation-oriented accounts are regular contenders in this ongoing competition, which comprises much of the literature on causation. Each of these definitions captures important characteristics of the notion of cause, but also raises difficulties that advocates of competing conceptions are quick to seize on. Needless to say, the contending accounts are never conclusively defeated by such difficulties; their advocates find ways to patch them as necessary. Nevertheless, the "attack/patch" cycle has an adverse cumulative impact as the difficulties pile up. More generally, concern over definitions and their weaknesses has led philosophers to devote a great deal of attention to intriguing yet marginal "hard cases." Adding "epicycles" may salvage a threatened definition of causation, but sheds little light on the ways in which causal notions are actually used, and in particular, on how they're used in scientific contexts. Science seeks to identify constraints that distinguish what may happen, or is bound to happen, from what is excluded from happening. Hence the notion of causal *constraint*, which is broader than the notion of cause, is at the center of my analysis. Even when searching for individual cause-events (effect-events), awareness of the framework of constraints that these individual events must satisfy is vital. And because there is no single causal constraint that is operative in science, but rather several different constraints, a study of the relationships between the various constraints is called for. I do not take the notion of cause to be reducible to any one of the constraints in question, or to a particular combination of them. The evolution of causal constraints—and thus of our understanding of causation—is as open-ended as the evolution of science in general. The difference between the "causal constraints" approach to causation, and the traditional approach, will become sharper as the book proceeds.

I will not review the current philosophical accounts of causation, and the difficulties they pose, in any detail. The *Oxford Handbook of Causation* (Beebee, Hitchcock, and Menzies 2009) gives an admirably balanced account of the literature. But it will be useful to briefly identify

the main contending proposals, and the key issues they bring to the fore, as these key issues underscore my claim that it is time for a different approach to causation.[2]

- **Regularity theories**, also known as Humean theories, reduce causation to lawful behavior and therefore assimilate causation to determinism.[3] Nonetheless, it is not laws that constitute causes, but the events that fall under them. Roughly, Hume's definition of causation set down three conditions: contiguity in space, succession, and constant conjunction of the same event types.[4] The succession condition can in turn be divided into a condition of contiguity in time, and an asymmetry condition to the effect that the cause must precede the effect. Physics has had to discard the spatial and temporal contiguity requirement due to its emptiness in continuous spacetime, and thus in mathematical theories that involve a continuum (such as theories employing differential and partial differential equations). But the remaining conditions are independent of the contiguity requirement and permit extension of the cause–effect relation to distant events: *any* event *c* that is regularly or lawfully followed by an event *e* can be considered the cause of *e*.

2. The few references in the following sections are only examples of relevant literature. Beebee, Hitchcock, and Menzies (2009) provides a wealth of bibliographic information.

3. The concept of determinism is touched on later in this chapter and examined in more detail in chapter 2.

4. Hume devotes extensive sections of *A Treatise of Human Nature* and *An Enquiry concerning Human Understanding* to causation. This brief summary ignores many interpretative issues and the vast literature thereon. Furthermore, it extends the meaning of "Humean" beyond anything Hume himself would have recognized. Even metaphysically extravagant accounts such as David Lewis's are often considered Humean, to say nothing of accounts, such as Mackie's and Davidson's, that deviate less radically from Hume's own account. Davidson's view is particularly interesting in this respect, as it is committed to the existence of laws that have no scientific utility. I would argue that Davidson's main point about causation (the distinction between causal and explanatory contexts; see below) is independent of the Humean commitment to laws. Other issues discussed under the rubric of "Humean causation" include the question of "Humean supervenience"—are there natural laws or only natural facts?—and questions about the status of laws—what sense, if any, can we give to the metaphor of laws "governing" the physical world? See, e.g., Maudlin (2007). I do not address these problems.

The connection Hume established between causation and lawful behavior has had a lasting impact on the philosophy of science. Yet regularity theories, though still among the leading accounts of causation, have also garnered much criticism. Objections target the very connection between regularity and causation, denying either the necessity or the sufficiency of regularity for causation. The claim that regularity is unnecessary for causation entails the acceptability of singular—that is, nonrepeatable—cause–effect relations. Although I do not deny the feasibility of such singular causal relations, their existence is peripheral to my primary concern. When focusing on science rather than, say, human actions, it is largely possible to remain within the boundaries of lawful causation.[5] The converse claim—that regularity is insufficient for causation—is backed by several arguments. For one thing, regularities, even when they appear to be lawlike, may reflect accidental rather than causal connections. For another, regularities as such lack the asymmetry typical of causal relations.[6] There are also examples like the tower and its shadow, which, despite the nonaccidental nature of the regularity in question, speak against the identification of lawful regularities with causation. The height of the tower and the length of the shadow are correlated by laws, but we see the shadow's length as caused (and explained) by the tower's height, and not the other way around. The example suggests a distinction between causal

5. The possibility of lawless causal relations and lawless actions is discussed near the end of chapter 7; the constraints examined up to that point are all lawful. I see no objection to considering singular events such as the Big Bang or the extinction of dinosaurs to be causes of later developments, but ascription of causal roles to singular events is not pivotal for my treatment of causation. There is an interesting exchange of letters, from 1920, between Einstein and Schlick on the question of causality without regularity. Einstein initially maintained that regularity was unnecessary, and suggested a hypothetical scenario in which such singular causal relations had to be posited. Schlick then convinced him that without regularities we could not even take measurements, and that our scientific notion of cause thus presupposes the existence of regularities. But Einstein continued to maintain that once we have formulated the concept of lawful causality, we should be able to identify singular causal relations (Albert Einstein Archive, 21: 576; Schlick 1920).

6. This asymmetry is touched on at the end of the chapter.

and noncausal regularities, where only the former are truly explanatory.[7] The concern that regularity falls short of causation often motivates the requirement that causal connections, unlike mere regularities, must be embodied in concrete mechanisms.[8] While the distinction between lawful and accidental regularities is crucial for science (and in that sense the critique of regularity theories is warranted), a connection between causation and concrete mechanisms is often lacking. When it comes to very general causal constraints, such as the relativistic limit on the speed of interaction, the search for an underlying mechanism is futile. The distinction between laws and mere regularities can be supplemented by a hierarchy that differentiates lower-order constraints, constraints on facts, from higher-order constraints, constraints on laws. Although constraints on laws do not fit the Humean scheme, they should be seen as causal constraints; see chapters 5 and 6. But even if we were able to fend off all the standard objections to the regularity account of causation, it would, from the perspective of this book, remain inadequate. Except for determinism, the constraints that make up the family of causal concepts (henceforth, causal family) cannot be expressed in the language of regular succession of individual events.

- The **counterfactual account** championed by David Lewis (1973) analyzes the causal relation between event *c* (the cause) and event *e* (the effect) in terms of the counterfactual "had *c* not occurred, *e* would not have occurred."[9] In addition to the formidable problem of analyzing counterfactuals,[10] and the

7. Arguing for a pragmatic approach to explanation, Van Fraassen (1980, 132) attempts to destabilize this intuition by depicting a case in which the length of the shadow is the motive behind construction of the tower. I would argue that even in this contrived case, it is the tower that causes and explains the shadow, but in any event, the example is not an instance of standard scientific explanation.

8. See Glennan (2009) and the literature cited there.

9. As Lewis notes, this account can be found in Hume's writings alongside the regularity account, giving rise to interpretative questions about Hume's "true" analysis of causation.

10. We will encounter some of these problems, in particular, sensitivity to description, in chapter 2.

metaphysical assumptions this analysis mandates, the counterfactual account faces the challenge of overdetermination. Recall the standard example (a typical hard, though marginal, case) of two desert travelers who set out, separately, to murder a third, one pouring poison into the victim's water flask, the other puncturing it, so that the victim will die of either poisoning or dehydration. The counterfactual conditional "Had *x* not punctured the water flask, *z* would not have died," fails to identify the cause, for (assuming the cause of death to be dehydration) it would not be true that had the water flask not been punctured, the victim would not have died. I should stress that I do not adduce these problems to critique counterfactual considerations in general, but to critique their adequacy as definitive criteria of causation.[11] I take counterfactuals to be indispensable for reasoning, and will use them extensively in chapter 2. But counterfactuals are also used in contexts that have nothing to do with causation: "If I were you, I would accept the offer"; "This triangle (pointing, for example, to a triangle with sides 3 cm, 4 cm, and 6 cm in length) is not right-angled—if it were, it would satisfy the Pythagorean theorem." Because of their broader applicability, counterfactuals cannot be relied on to pick out causal relations.

- **Process accounts** of causation focus on prolonged progressions rather than instantaneous events, and tie causation to a particular process, such as energy transfer from one system or state to another (Fair 1979; Salmon 1984; Dowe 1992, 2000). This approach can handle the case of Jane's happiness being due to John's response to her message but has difficulty with the seemingly parallel case of Jane's unhappiness being due to John's failure to respond. In other words, the process approach is unable to account for failures and omissions.
- **Probabilistic accounts** of causation (Suppes 1970; Kvart 1986) have the great advantage of extending causal discourse to non-

11. Critique of counterfactual reasoning is quite common; see, e.g., James ([1884] 1956).

deterministic contexts. From the probabilistic perspective, a cause need only raise the probability of the effect-event's occurring; there is no need for it to determine or induce its occurrence. On the other hand, however, probabilistic accounts engender paradoxes of their own—namely, probabilistic correlations that do not seem to reflect causal relations, or events that seem entitled, from the intuitive point of view, to be considered causes of certain "effect" events, yet appear to lower, rather than raise, the probability of their occurrence.

- **Interventionist or manipulation accounts** of causation (also known as agency accounts) have been in vogue for several decades (Menzies and Price 1993; Woodward 2003; Hitchcock 2007). Here, causes are identified as the factors that, when manipulated, change the result that would have ensued in the absence of that intervention. In identifying causes as necessary conditions of effects, the manipulation account has much in common with the counterfactual account. In focusing on human intervention, however, it has given rise to the objections of anthropocentrism, and—since intervention is actually a causal concept—circularity. One merit of the interventionist criterion is that it distinguishes causal relations from correlations that are merely accidental. Together with insights from the counterfactual and probabilistic accounts, it has stimulated elegant work on causal networks and their graphic representations (Pearl 2000). Causal networks are highly valuable in a variety of practical contexts—legal, medical, economic, policy-making, and so on—where distinguishing effective from ineffective intervention is essential. Nevertheless, the manipulation account, I contend, is far too limited to provide a comprehensive understanding of causal processes in the world. This point requires elaboration.

Consider, first, the contrast between the regularity account and the manipulation account. In a standard case, where natural laws and initial conditions determine a certain result (a certain chemical reaction, say), we can often manipulate the initial

conditions, but not the laws. On the manipulation account, therefore, the initial conditions constitute the only relevant factor for a causal account of the process. By contrast, the regularity theorist ascribes a crucial role to the laws—to what we *cannot* manipulate.[12] The initial conditions can be considered causes only because they are invariably followed by the same trajectories, that is, they are considered causes only because of the existence of laws that are non-manipulable constraints. And although (for the reasons mentioned above) I do not consider the regularity account, as it stands, fully satisfactory, there is something fundamentally correct about the intuition that constraints which we cannot manipulate are an inherent feature of causal descriptions and explanations. Furthermore, invoking laws is by no means the only context in which we ascribe a causal role to the non-manipulable. The electric charge of the electron, which though non-manipulable is causally efficacious, is a case in point.[13] The constraints we will examine in this book are typically not subject to human intervention, but they enable us to grasp and predict the dynamics of unfolding events, and exclude infinitely many alternatives to what actually transpires. In this sense, they also constrain our interventions—manipulations and interventions are always carried out *within* a general framework of constraints. The manipulation account of causation thus already *presupposes* the preconditions of possible manipulation, preconditions that a complete causal story must take into account and render explicit.

12. Manipulation theorists such as Woodward (2003) make it clear they do not restrict manipulation to procedures and actions that humans can actually carry out, but allow for a broader class of manipulations that are possible in principle. Laws of nature, however, are beyond control even in this weaker sense.

13. A more controversial example is Newtonian space, which plays a causal role in Newton's theory—acceleration relative to absolute space has genuine physical effects—yet this space and its geometric structure are fixed and cannot be manipulated. This interpretation is in line with Einstein's view of the matter (see Mutuality of Causal Relations at the end of this chapter). See, however, DiSalle (1995) for a critique of this causal interpretation of space.

In view of the difficulties that beset each of these accounts, I have relinquished the search for *the* definition of causation, instead taking causation to be a cluster concept comprising a broad range of causal notions. My primary focus will be the causal notions employed in science, which include the much-discussed notion of determinism, but also notions such as stability and locality, which philosophers tend to neglect. In turning to pluralism, I am not, of course, alone: a pluralistic attitude to causation has been advocated (sometimes only in passing) by Reichenbach (1956); Anscombe (1971); Cartwright (1983, 2004); and Godfrey-Smith (2009), not to mention Aristotle, who introduces his presentation of the four causal categories by noting that the number of causes matches that of the things "comprehended under the question 'why' " (*Physics* II, 198a15–16). Skyrms (1984) speaks aptly of an "amiable jumble" of causal notions that can, but need not, work together.

But the declaration of pluralism is only a starting point. To do justice to causation, recognition of the variety of causal notions must be augmented with detailed investigation of their usage, especially in fundamental science. Science, like daily life, presents us with a spectrum of causal notions and constraints. These scientific constraints are often in some way "descended" from the more intuitive constraints of daily life, though differing from them significantly in precision and scope. The invocation of causal notions in scientific contexts is particularly noteworthy in view of arguments that challenge causation's place in fundamental science, or relegate it to "folk science" (Russell 1913; Norton 2007). I will return to the Russell-Norton position further on, but for now, let me point out that the argument I make in this book is that as soon as we shift our attention from the familiar paradigms of breaking a glass or tickling a baby to determinism, locality, stability, and conservation laws, it becomes evident that causal notions permeate fundamental science.

Critique of causal discourse also comes from another direction. These critics concede causation's place in fundamental physics, but deny it elsewhere, arguing that higher-level realms, such as the realms of biological, mental, and social events, are causally inert. Alleged causal relations on these levels, or between higher-level events and events at

the fundamental level, are (according to this view) all reducible to the causal relations of physics. This argument is rebutted in chapter 7.

Our first encounter with causal notions does not come from fundamental science. We acquire the notion of cause early in life in relatively simple situations, such as sucking, pulling, pushing, holding, biting, and so on. These actions involve several elements of the cluster concept of cause, so that many of the features explicated in the aforementioned accounts of causation are operative. A child pulls the string of a toy that plays a tune. The interaction is both regular—whenever the string is pulled the tune plays—and local—no action at a distance; it instantiates both the manipulative and the counterfactual conditional accounts of causation—had the string not been pulled, no tune would have been played; and there is no creation ex nihilo—the power was provided by tugging the string, energy was transferred from the child's hand to the string, and from the string to the musical instrument. Infants are unable to articulate these concepts, but may acquire a rudimentary grasp of some of them, and learn to associate the features in question with each other, so as to form a more complex sense of causation. Later on, with exposure to less paradigmatic causal nexuses, and to science, intuitive ideas give way to more explicit notions, occasionally undoing the automatic associations, or establishing new ones. A child who is at one stage prone to "magical" thinking, for instance, believing that merely wishing someone ill suffices to actually bring about harm, will, with age and experience, likely revise this conception.

A recurrent concern about causality, dating back to Hume, derives from empiricism: do we ever *observe* causal connections? And if not, ought we not either renounce causation or reduce it to observable features of the world? Anscombe's influential *Causality and Determination* (1971) argues, contra the Humeans, that we do indeed observe and experience numerous *instances* of causal connection: pushing, breaking, burning, and so on. I agree with Anscombe. Granted, there are also many less evident cases, where the causal connection is not observable, but the same goes for other relations; they too are manifest in paradigmatic situations and remote from immediate experience in others. Motherhood, for instance, is not in general a directly observable rela-

tion, but when we happen to be present during a delivery, we can witness it directly. And we can now adduce empirical evidence of motherhood in DNA sequences. This evidence does not render motherhood directly observable, but establishes it beyond a reasonable doubt, making it as objective as other empirical relations. Moreover, it has long been acknowledged that science does not restrict itself to the directly observable; it is empirical only in the sense that it expects its nonobservable concepts and laws to have observable implications. In this respect, causal thinking is in line with scientific thinking in general. My perspective on causation is realist, and I take causal constraints to be objective. This realism should not be construed as a commitment to the existence of causes as metaphysical entities that exist in addition to the entities they relate. Even for those who do not embrace his account in its entirety, Hume's critique of the traditional conception of causation has discredited this picture of causes as hidden "arrows" between events. The characterization of realism in mathematics as a commitment to the objectivity of statements rather than the existence of mathematical objects (Kreisel 1958) can be applied, mutatis mutandis, in the context of causation: realism about causation means that causal claims are objectively true or false.

I should, perhaps, note that the empiricist status of causation has undergone an ironic transformation. Hume deemed spatial and temporal relations legitimate from the empiricist point of view, using them as the basis for his definition of causation in terms of constant conjunction. But the relationship between the causal and temporal orders turned out to be quite different from that which he envisaged. According to the special theory of relativity (STR), the temporal relations between events are only well defined in regions of spacetime charted by light signals representing (and limiting) the *possibility of causal interaction*. When events are separated by space-like distances, there can be no causal interaction between them, and consequently, their temporal order is not invariant, but varies with the coordinate system. Rather than being reducible to spatiotemporal relations, causality now appears to be the basis for the very structure of spacetime. Causal relations are thus at least as fundamental as temporal relations, and arguably

(as suggested, for example, in Reichenbach 1956), conceptually prior to temporal relations.

Typically, definitions allow us to replace the defined term (*definiendum*) with the terms that define it (*definiens*). When, for instance, Mackie (1965) argues that a cause is (at least) an INUS condition, he is suggesting that "the short-circuit caused the fire" could be replaced with "the short-circuit was an INUS condition of the fire."[14] By contrast, causal *constraints*, being necessary but insufficient conditions, do not replace the notion of cause in this way. Even so, they constitute an essential aspect of using and understanding causal discourse. In accepting Dan's alibi to the effect that at the time his Cornwall cottage was set on fire, he was in Oxford giving a talk, the court takes it for granted that there is no action at a distance. Although this causal assumption constrains our identification of causal connections, it does not allow us to replace locutions signifying causal constraints with the term *cause*, nor does it identify the cause of the fire. Similarly, when physicists refer to the limit on the speed of interaction as "relativistic causality," they are using the term *causality* to refer to a constraint—a necessary condition—that neither defines causality nor points to the cause of a particular process or event. The same caveat applies to other constraints, such as symmetries, conservation laws, and variation principles: they circumscribe our causal thinking but do not provide us with synonyms for the term *cause* or coextensional alternative locutions. It appears that no one constraint constitutes a condition that is both necessary and sufficient and could thus serve as an adequate definition of the notion of cause, a definition that covers all its applications. Hence there is also no general "causal principle" (more on this later). I have, therefore, relinquished the quest for such a definition and adopted a pluralistic approach, taking the notion of cause to be an irreducible cluster concept covering various constraints imposed by the theories we employ. The cluster's com-

14. An INUS condition is a condition that in itself is neither necessary nor sufficient for the occurrence of the effect, but constitutes a necessary component of a complex that is sufficient but unnecessary for bringing about the effect. INUS stands for Insufficient but Necessary (in a cluster that is) Unnecessary but Sufficient.

ponent concepts do not apply across the board; where they apply, and to what level of precision, is an empirical question.

The fact that the causal constraints examined in this book do not presuppose individual cause-events or effect-events, or that such putative events cannot always be picked out, is no drawback: in scientific contexts, cause–effect language (and ontology) is not always helpful. Indeed, it is generally cause–effect ontology, as opposed to causal discourse per se, that is targeted by critiques of the concept of cause. We can understand why a certain chemical reaction occurs in the direction it does by pointing to symmetry principles and conservation laws. From the perspective adopted here, this would be a perfectly adequate causal explanation, though it does not single out any individual event as the cause of the outcome. I am not claiming that we can always dispense with identification of individual cause-events, or that we ought to do so.[15] Nor do I want to question the utility of the ontology of events: we certainly do wish to apply the general constraints to individual systems, states, and events. But typically, these applications do not pinpoint any single event as a cause, and all the more so, as *the* cause. There are contexts, however, not only in daily discourse, but also in science, in which ascription (or denial) of causal roles to individual events becomes essential. In assessing the implications of relativistic causality, for example, it is sometimes crucial (as we will see in chapter 4) to identify causal relations and the transmission of information (or lack thereof) between individual events. But even when such ascription of a causal role to specific events is irrelevant or meaningless (as in the example of Newtonian absolute space), it does not follow that there is no causal story to tell.

In the literature, considerable effort is devoted to distinguishing causes from conditions, and singling out *the* cause of an event from other events that stand in a causal relation to it but lack some feature that would make them the sole cause. Speeding might be singled out as the cause of an accident, while the curve in the road, the weather, or the design of the vehicle, are deemed mere "conditions." The distinction is

15. I do not deny the existence of causality in singular cases, see note 5 earlier in this chapter.

generally considered pragmatic and thus context-dependent. The idea is that whereas, from the logical point of view, many factors stand in a causal relation proper with the effect-event, from the pragmatic point of view, it is legitimate to distinguish just one of them as its cause. Pragmatic criteria include, for example, deviation from the regular course of events, and human intervention (Hart and Honoré 1959; Mackie 1965). Both these criteria distinguish the speeding from the other conditions that played a causal role in bringing about the accident. But in a different context, say that of evaluating the road's safety, the curve in the road might be the focus of blame and the target of intervention, whereas speeding drivers are deemed background conditions. Such pragmatic considerations are not foreign to scientists, who invoke them routinely when planning experiments and analyzing their results, but with regard to the causal constraints addressed in this book, they can be set aside.

What, then, is the role of causal notions in science? Causal notions and constraints, I suggest, are employed to describe, predict, and explain change. They tell us which processes and changes in the physical world are possible, and which are not. This characterization gives us a far broader picture of causation than the picture painted by portraying only cause–effect relations. Causal notions in this broader sense, though not the only explanatory notions, are unique in explaining *change*. Logical and mathematical notions may also play an explanatory role, and, in the realm of human action, reasons, arguably different from causes, fulfill a central explanatory function. But our understanding of change in the physical world is not, and cannot be, complete without causal notions. Typically, causal notions involve changes that occur over time, a characteristic that distinguishes them from the nontemporal relations found in mathematics. And they also involve matter—masses, forces, fields, and their interactions—which again sets them apart from the purely mathematical. Thus even when expressed in mathematical language, causal relations and constraints go beyond purely mathematical constraints; they are (at least part of) what we add to mathematics to get physics.

Causal relations in the physical world have not always been properly distinguished from necessary connections in the logical-mathematical

realm. In Spinoza's system, for instance, logical necessity and causal necessity are on a par. But the disentanglement of these kinds of necessity has become common, if not mandatory, in modern science. Time, matter, and the possibility of change are crucial to maintaining a distinction between physics and mathematics. To reiterate my characterization, a causal constraint is any constraint that delimits change, distinguishing changes that are sanctioned by science from those that are ruled out. The test of legitimacy is empirical: instances of legitimate change are detectable in the physical world, excluded changes are not. This conception of the difference between physics and mathematics generates a natural account of causation in terms of temporal change and the constraints that such change must satisfy. One alternative to this natural account is to erase the distinction between physics and mathematics altogether and embrace a hyperrationalist picture of the world, such as Spinozism, or a hypermathematical one, such as Pythagoreanism. I find this alternative unappealing.

Lange (2017) argues for another alternative.[16] In addition to causal explanations, on the one hand, and purely mathematical explanations, on the other, he introduces a third category of explanations, which are neither causal nor mathematical. Interestingly, he focuses on the notion of constraint in this context, referring to this third category as "explanation by constraint." Lange's attentiveness to constraints is commendable, but whereas the constraints I speak of are *causal*, he deems explanation by constraint noncausal, implying that the notions of cause and constraint exclude each other. The rationale for this exclusion seems to be recognition of a hierarchy of laws, some being more general—and hence, in Lange's view, more necessary—than others. Lange takes Newton's second law of motion to be higher up in this hierarchy than the law of universal gravitation or Coulomb's law, because Newton's second law applies to forces in general and would presumably apply to new forces, were any to be discovered. The hierarchical picture is apt—some causal constraints do indeed apply to *laws* rather than events—but why reserve causal status for lower-level laws? Imposing this restriction

16. I thank an anonymous reader for calling my attention to this recent book.

involves Lange in extended discussions of what does and doesn't count as causal explanation, the sort of analysis that scientists do not engage in, and I am trying to avoid.[17] The inclusive concept of causation as comprising *any* constraint on change, regardless of its place in the hierarchy of laws, affords a better understanding of the function of causal notions in science.

Chapters 5 and 6 provide illustrations of the causal role of higher-level principles such as symmetries and variation principles, and of the difference between mathematical and physical constraints. Symmetries, for instance, are expressed in mathematical language—the language of group theory—which gives them a formal, even a priori, appearance. But there are facts about the world that, despite their expression in this group-theoretical language, cannot be considered mathematical facts. A physical process may be invariant under spatial translation, a feature that is reflected in the mathematical formulation of the process and the laws it obeys, a formulation that is independent of specific coordinates. But the fact that this is the correct formulation of the law, viz., that this symmetry is reflected in reality, is not a mathematical fact—we could envisage physical processes and laws that are not invariant in this way (as was actually the case in Aristotelian physics). Furthermore, among the symmetry considerations adduced by physicists, we find Curie's principle, according to which (roughly) we cannot get asymmetry from symmetry. Rather than being a purely mathematical theorem, Curie's principle (discussed in chapter 5) identifies changes we can expect to find in the physical world; that is to say, it is a causal principle. The causal constraints encountered in fundamental science go well beyond the intuitive causal concepts we grew up with, and well beyond the examples that recur in the philosophical literature on causation. Inclusion of symmetry considerations in the causal family is a prime example of the extension that is called for when we move from causal relations between

17. E.g., Lange elaborates on a distinction between cases in which the law of conservation of energy functions as a causal explanation, and cases in which it functions as an explanation by constraint (2017, chap. 2). I see both cases as clear instances of causal explanation. Note also the contrast between his account of Pauli's exclusion principle (183) and my causal account of this principle in chapter 5.

individual events to a much more general understanding of physical change.

It might be objected that temporality no longer distinguishes physics from mathematics; the time required for a calculation, for instance, is a major parameter in computation theory. But this objection is misguided. The temporal terminology in computational science is premised on the realization that computations are carried out by physical systems—human beings or machines—that are constrained by physical possibility and cannot perform instantaneous calculations. But this realization is not part of mathematics. The fundamental notion of computational theory—the number of steps required—is indeed mathematical. But though the length of a calculation in terms of the number of steps it takes is a mathematical consideration, its length in terms of the time it takes is not—it involves assumptions about the physical world. Similarly, the causal properties ascribed to algorithms such as those of cellular automata are also a figure of speech. In describing John Conway's Game of Life (Berlekamp, Conway, and Guy 1982; Gardner 1970), we might say, for example, that step n is the cause and step $n + 1$ the effect, but this formulation tends to conflate the algorithm with the computer that implements it. The computer does indeed operate in a causal manner, using electric circuits and the like, so that each of its states is causally related to earlier ones. The algorithm, however, is temporal and causal only in a metaphorical sense.

To fully understand change, we must be able to understand not just what happens, but also what *fails* to happen or is *excluded* from happening. By the same token, it is just as causally relevant to learn that a system is *in*sensitive to a certain parameter as to learn that it is sensitive to it. Traditional accounts of causation do not fully address this negative aspect of change and causation. It should be noted, first, that there are different kinds of exclusion. When we think of an event c as bringing about an effect e, it is implied that whereas e must, given c, occur, every other outcome is excluded. This kind of exclusion is specific to a particular state of affairs; an event that is *ruled out* under one set of initial conditions may be permitted, or even mandatory, under different initial conditions. There are, however, types of events—certain chemical or

nuclear reactions—that *never* occur, regardless of the initial conditions. From the physicist's perspective, what does not happen has just as much causal significance as what does. Regarding such absolute exclusions, the working assumption is that there must be some underlying principle that explains them. Symmetry principles and conservation laws provide explanations of this kind and are sometimes explicitly formulated in terms of exclusion rather than affirmatively, in terms of what they mandate. Pauli's exclusion principle (discussed in detail in chapter 5) is a case in point. As noted, it is also important to distinguish between constraints that bar specific event types and constraints on the general form of laws. Symmetries exemplify the latter type of constraint as well. They are therefore often considered to rank higher in the hierarchy of laws of science than ordinary laws.

We can think of physical constraints on legitimate change in terms of an analogy that invokes two different models of legality. On the first, usually deemed applicable to state officials, everything that they are not mandated by law to do is prohibited.[18] On the second, usually deemed applicable to citizens, everything that is not prohibited by law is permitted. Similarly, we can think of the constraints imposed on natural processes either as necessitating everything that happens, or as excluding certain occurrences, but leaving a considerable amount of freedom: whatever is not excluded may happen. On the former model, science is expected to show that the occurrence of anything that happens is determined by law, while the occurrence of anything else is excluded by law. This expectation places severe restrictions on what will be considered an adequate scientific theory. On the latter, freedom-granting model, science is only expected to show compatibility with the law. That is, it suffices for science to formulate laws that are not violated, or to put it differently, laws that permit, rather than determine, what happens. In principle, the two models could converge. That is, it could be the case

18. The analogy is incomplete: first, because in the legal realm laws are normative and are often violated in practice, whereas in science they are descriptive and, to the extent that they are true, cannot be violated. Second, the freedom-excluding model is too extreme; officials, even in their capacity as such, have some liberties. Overall, however, for officials, the freedom-excluding model is the default account.

that we begin by listing the proscriptions, with the intent to delimit the freedoms that remain, but by the time we have the complete list of proscriptions, we discover that there is no room left for freedom, and everything that does happen is in fact determined by law. Many scientists strive to demonstrate the reality of this rigid scenario, seeking to make the list of exclusions sufficiently comprehensive to eliminate any freedom whatsoever. Such an ambition was voiced by Einstein, who pondered the question of whether God had any choice when creating the world. On the no-choice picture, brute facts, contingencies that are inexplicable by fundamental laws, are an embarrassment to science. The theory of the Higgs field, and the search for the Higgs boson that confirms it, are motivated by a desire to derive particles' masses from general principles rather than accept their values as contingent and inexplicable parameters.

The freedom-excluding scenario is not, in fact, realized in contemporary physics, where the two models coexist. The freedom-excluding model has obvious links to determinism, but recall that even when laws are deterministic, the question of freedom might still be applicable to the initial conditions.[19] On the other hand, quantum mechanics (QM), its indeterminism notwithstanding, invokes more symmetry principles than are recognized in classical mechanics, taking them to be strictly (rather than probabilistically) obeyed.[20] As a rule, symmetry principles fit the freedom-granting model; they exclude certain processes, certain nuclear reactions, say, but leave room for more than a single possible outcome. A particularly interesting combination of freedom and necessity can be found in Feynman's picture of quantum mechanics, where freely moving particles seem, nonetheless, to behave as if they were exemplifying the rule that "everything that can happen does happen."[21]

19. Newton, e.g., thought that the solar system's initial conditions were not determined by the laws of physics, but ensued from God's benevolent choice; these initial conditions were later derived from mechanics. The general question about initial conditions, however, is still open, and is often a matter of controversy, as in statistical mechanics. See Albert (2011).

20. For the moment, I ignore approximate symmetries.

21. The question of the relation between this tenet and the traditional causal principle(s) merits examination, but I will not take it up here.

Cox and Forshaw (2011) adduce this tenet as the theme of their Feynmanian exposition of quantum mechanics. A similar view was debated in the seventeenth century, though with the theological gloss prevalent at the time. Leibniz, in disparaging this view, argued that to expect God to realize every possibility, regardless of its merit, is comparable to the expectation that a poet should "produce all possible verses, good and bad" (Strickland 2006, 137). Feynman's version of QM, and its implications for causation, are discussed in chapter 6.

Up to this point I have given two reasons for broadening our conception of causation beyond its familiar philosophical habitat. First, the causal notions and constraints explored here are all required for a comprehensive understanding of changes that take place in the world, and are the tools scientists employ to acquire such an understanding. Second, omissions and exclusions, which are integral to any account of causation in science, but constitute notorious stumbling blocks for most philosophical accounts of causation, fit smoothly into the picture suggested here. There are two further considerations that support the broader approach to causation. One is historical: the causal constraints of contemporary science are the progeny of intuitions and assumptions that have been associated with causation for as long as memory serves, and are therefore rooted in a long tradition of causal discourse. The other is conceptual, and pertains to the links between different constraints. Viewing causality as a manifold enables me to bring to the fore questions about the relationships between determinism and locality, determinism and stability, stability and symmetry principles, stability and variation principles, and so on. These questions, which have received little philosophical scrutiny, can be tackled from a general conceptual viewpoint or from the perspective of a particular scientific theory. Such an investigation will yield answers that support my reading of causation as a family of interrelated concepts. But let me add that these interrelations are worthy of scholarly attention regardless of the validity of my claim that all the constraints in question are in fact members of the causal family, and are needed for explication of the notion of causation.

I have been moving freely between the context of causation and that of causal explanation as if they were interchangeable. To be more pre-

cise, we should follow Davidson in recognizing a crucial difference between the two contexts. In his celebrated "Causal Relations" ([1967] 1980), Davidson observes that while truth values of singular causal statements are independent of the descriptions used to refer to the related events, explanatory contexts, much like other intensional contexts, are description sensitive. Compare, first:

1. Lord Kelvin made significant contributions to thermodynamics.
2. Sharon knows that Lord Kelvin made significant contributions to thermodynamics.

The truth value of (2), unlike that of (1), may change when "William Thomson" is substituted for "Lord Kelvin," for Sharon may not know that William Thomson is Lord Kelvin. Davidson points to an analogous difference between (3) and (4):

3. The reaction caused the explosion.
4. The reaction explains the explosion.[22]

According to Davidson, singular causal relations are extensional—the truth values of sentences affirming (denying) them do not change when we refer to the same entities by means of different descriptions. By contrast, explanatory contexts, being sensitive to the descriptions of the events in question, are referentially opaque. This opacity is due to the fact that explanations comprise laws that connect *types* of events rather than individual events. To explain an event by subsuming it under a law (or set of laws), we must, therefore, refer to it by means of the right description—namely, the description matching the event type specified by the relevant law(s). Davidson's insight has often been overlooked, but is crucial for a proper understanding of the notions of determinism and stability. As we will see in chapter 2, Davidson's point is particularly relevant for the assessment of Russell's critique of causation qua determinism. Moreover, description sensitivity is characteristic of probabilistic explanations as well. Statistical mechanics provides an

22. My examples differ slightly from Davidson's. (3) and (4), in particular, are short for what Davidson formulates as: "That the reaction occurred caused it to be the case (explains) that. . . ." Sidney Morgenbesser is known to have made the same point. See also Steiner (1986).

instructive example of the scientific significance of descriptive categories. Here, the fact that macrostates vary enormously in the number of microstates they comprise is of crucial explanatory importance. Since we ourselves define macrostates, the explanatory import of statistical mechanics hinges on description-sensitive facts. Description sensitivity, however, does not breed subjectivity, as has been alleged. Once a description is chosen, the claims made in terms of this description can be objectively true or false. There are, of course, natural and unnatural, useful and useless descriptions, and finding the most helpful descriptions is far from trivial. But these obstacles do not entail any conflict between description sensitivity and objectivity. These points are elaborated on in chapters 2 and 3.

As mentioned, the role of causation in science has been a matter of controversy. Russell (1913) dismissed causation, arguing that mature sciences, physics in particular, consist of differential equations that do not invoke the notions of cause and effect.[23] More recently, John Norton (2007) has revived this negative attitude, arguing that the notion of cause can be tolerated in "folk science," but not in fundamental science. This position is referred to as the "eliminativist" or "error" theory of causation. Alluding to Russell's famous remark that the law of causality survives, "like the monarchy, only because it is erroneously supposed to do no harm" (1913, 1), Norton and other members of the dismissive camp are sometimes also referred to as "republicans" (Price and Corry 2007). Norton draws an analogy to science, where advanced theories typically recover the results of less advanced theories in some limited way—the predictions of classical mechanics, say, are derived from those of the special theory of relativity for velocities much lower than that of light. He thus seeks to recover the causal structure of our commonsense picture of the world from the more accurate depiction generated by fundamental science, which he takes to be completely free of causal notions. Despite their concurrence vis-à-vis causality, Russell's objection to determinism is quite different from Norton's: whereas

23. Quine too noted that "the notion of cause itself has no firm place in science" (1966, 229). Interestingly, Russell (1948) rehabilitates causality, espousing a view that is closer to process accounts than to regularity accounts.

Russell maintains that determinism is empty and trivially satisfiable by any theory, Norton argues that determinism is *false* even in the context of the theory considered its safest harbor—classical mechanics. I discuss Russell's position in chapter 2. Note, however, that in their critique of causation, both Russell and Norton actually focus on *determinism*, which is obviously a narrower concept of causation than that which I am recommending here. Their arguments, even if accepted, leave intact other causal constraints' applicability and usefulness in fundamental science.

The republican critique actually has two targets: the notion of cause and the causal principle. Although these concerns are not identical, both Russell and Norton connect them. Russell, as we saw, derides the causal principle, but also claims that "the word 'cause' is so inextricably bound up with misleading associations as to make its complete extrusion from the philosophical vocabulary desirable" (Russell 1913, 1). Norton links the notion and the principle even more directly: "Centuries of failed attempts to formulate a principle of causality, robustly true under the introduction of new scientific theories, have left the notion of causation so plastic that virtually any new science can be made to conform to it" (Norton 2007, 12). I grant this premise—there may be no "principle of causality" whose truth is secured a priori or established beyond reasonable doubt by experience. Indeed, there is not even an agreed-on formulation of the traditional principle. But the redundancy of the concept of causation does not follow from the demise of the causal principle. (Would it not be an overreaction to give up the *concept* of justice just because we are unable to formulate an overarching *principle* of justice?) Combining critique of the principle of causality with critique of the concept of cause (as Russell and Norton do) is, perhaps, understandable if one identifies causality with determinism and takes determinism to imply a very general principle about reality, such as "every event has a cause."[24] The failure of this general principle is then taken to imply the futility of the very concept of causation. But in my view, the principle as an assertion about the world (or our best theory of the

24. For the moment, this ancient version of the causal principle will do; more accurate formulations follow below and in chapter 2.

world) should still be distinguished from the concept. After all, it could be the case that some systems or processes obey deterministic laws though others do not, in which case the concept would be applicable despite the fact that the general principle fails. Thus, from the broader perspective adopted in this book, an open-minded attitude to the principle is appropriate. Rather than aspiring to a consensus regarding a universal causal principle, we must make do with a family of causal constraints that, like other natural laws, are subject to repeated testing and refinement.[25] And the same goes for the family of causal notions—they too must prove their value to science through their scientific applications.

In more recent publications (Frisch 2009a, 2009b; Norton 2009), the controversy over causation in science has shifted from the question of whether there is a meaningful principle of causality to the question of whether there is an *independent* principle of causality, that is, a principle leading to results that could not have been reached by any other physical principle. But if, as I maintain, conservation laws and variation principles *are* causal principles, then any result derived from them—and such results abound in physics—is derived from a causal principle or causal constraint (even if not from the sort of singular causal principle that Russell and Norton are so dismissive of). Moreover, seeking to establish the existence of an "independent" principle of causality is uncalled for. We might just as well ask whether there is an independent notion of a family, that is, if there is a family relation over and above, and independent of, being a daughter, brother-in-law, cousin, and so forth. Clearly, family relations can be subdivided, but does this make the notion of family redundant? I would argue that it doesn't, but the status of the general concept (of family and cause alike) is not the main issue. There may be no significant difference between the two pictures of causation—a single concept made up of several components, and a cluster of distinct concepts that are closely interrelated. If so, the debate over the term *causation* dissipates into a minor verbal disagreement. I want to stress, however—and this goes beyond the merely verbal—that gen-

25. Some of these constraints are quite general; see the discussion of Curie's principle in chapter 5.

uine questions remain about the relations between the various subconcepts. Regardless of whether we deem the notion of family redundant, we should be able to answer the question of whether cousins, say, can also be brothers. Analogously, regardless of whether we deem the notion of causation redundant, we should be able to answer the question of whether determinism implies locality or stability. This book is written from a pro-causation perspective, but the project it tackles—analysis of the relationships between members of the causal family—should, I believe, engage "republicans" as well.

To familiarize ourselves with the causal family, let me briefly introduce some of its members, emphasizing their connections to earlier traditions and intuitions about causation.[26]

DETERMINISM. The most prominent member of the causal family, determinism is frequently taken to be the core meaning of causation. It is also the meaning most closely associated with the so-called causal principle. Although the term *determinism* was coined in the nineteenth century, the ideas associated with determinism, such as exclusion of chance, go back to antiquity, and have been widely discussed ever since, under a range of rubrics, in particular causality and necessity. Determinism calls to mind two earlier principles: the universality principle, according to which nothing happens without a cause, and the regularity principle, according to which the same (type of) cause invariably leads to the same (type of) effect. In themselves, these principles are neither equivalent nor coextensional—a world satisfying one of them can violate the other. If, however, regularity is considered constitutive of causality (as it is in Hume's analysis), then a world satisfying the universality principle also satisfies the regularity principle. The converse does not follow. Despite the fact that "determinism" is often invoked as a feature of reality (for example, when debating the problem of human freedom), it is preferable to think of it as a property of theories. On the contemporary understanding, a theory is deterministic when it implies (roughly) that the entire trajectory of a closed physical

26. Clearly, specific laws such as Newton's laws can also be thought of as causal constraints, but I think of the causal family as including general rather than specific constraints.

system is determined by its initial conditions (or indeed, its conditions at any particular moment). When this stipulation is met, both regularity and universality obtain.

The contemporary definition of determinism thus combines the features that were traditionally thought to characterize the causal nexus, thereby linking causation and determinism. The differential equations of theoretical physics highlight this connection. Einstein put it as strongly as this: "The differential law is the only form which completely satisfies the modern physicist's demand for causality" ([1927] 1954, 255). Surprisingly, though, in restricting itself to closed systems, the contemporary definition of determinism creates new problems for some accounts of causation. By definition, a closed system cannot be interfered with. If one conceives of causality along the lines of the manipulation account, then, as Stachel (1969) has convincingly argued, determinism and causation are incompatible; the former can only be satisfied in closed systems, the latter in open ones. From the perspective of this book, however, neither the identification of determinism with causation, nor the claim that they are incompatible, is justified. Determinism is but one type of causal constraint, one member of the causal family. Its subtle relations with other constraints will be explored in detail in the coming chapters.

LOCALITY. Although in many contexts, the concept of determinism is taken to be synonymous with that of causality, there are also contexts—in particular, the context of the special theory of relativity (STR) and its relation to quantum mechanics (QM)—where it is the term *locality* that is typically used interchangeably with *causality* (or *relativistic causality*). A descendant of the traditional "no action at a distance" constraint, as well as the earlier *Natura non facit saltum* principle, locality is a constraint that excludes spatial or temporal gaps in physical interaction. The idea underlying the term *locality* is that changes in the physical world follow local "instructions" from the immediate environment rather than instantaneous ones from distant locations. Satisfying the desideratum of locality is one of STR's advantages over Newtonian mechanics, which involved instantaneous gravitational interaction between distant masses. The fact that both *determinism* and *locality* are

used interchangeably with *causality* may lead us to assume that these terms are closely related, or at least coextensional. As we will see in chapter 4, however, locality and determinism are distinct concepts that figure in various intricate relations in different theories. Their interrelation is particularly intriguing in the framework of QM, where entangled states exhibit nonlocal correlations that have been alleged to pose a threat to QM's compatibility with STR. To preempt this threat, the relativistic constraint of locality has been narrowed down to no-signaling. That is, nonlocal correlations are legitimate as long as they do not allow the transmission of information between distant (though correlated) events. We will see in chapter 4 that indeterminism is the key to peaceful coexistence between QM and STR.

STABILITY. A stable state is a state to which a system tends to return after having been slightly perturbed. Stability might be the phenomenon we seek to explain: explaining the stability of atoms, for example, was one of the problems that led to the discovery of quantum mechanics. But stability is also an important explanatory notion adduced to understand the prevalence of one type of state, say equilibrium, over another type of state known to be erratic or short-lived. Unlike determinism and locality, the notion of stability does not rest on classical intuitions about causation. This might reflect the fact that, despite its explanatory import, stability does not constitute a general causal constraint. Depending on various factors, such as the nature of the relevant boundary conditions and the kind of perturbation involved, the same laws are compatible with the existence of both stable and less stable states. A system obeying deterministic laws can thus reach stable or unstable states, and the same is true of stochastic systems. Stability must therefore be carefully distinguished from determinism. In chapter 2, I argue that the conflation of these concepts, which is not uncommon, leads to serious blunders, and in particular, to imputing teleology to nonteleological processes. A better understanding of the notion of stability can serve to obviate teleology in a variety of contexts: history, evolutionary theory, mechanics, and statistical mechanics. The terminology used in these contexts may differ from that used in physics. Analysis of the concept of stability will therefore be accompanied by explication of

related notions such as necessity, contingency, robustness, and resilience, all of which suffer from vagueness and ambiguity. The notion of stability is also invoked to elucidate the relationships between different physical levels, quantum and classical mechanics, classical and statistical mechanics. Exploration of the concept of stability is thus edifying vis-à-vis debates over reduction and emergence, examined in chapter 7.

CONSERVATION LAWS. That some physical quantities are conserved, whereas others are not, can explain why certain interactions are commonly observed, and others, never encountered. Like determinism and locality, conservation laws are constraints on possible change, and as such, they articulate our understanding of causation. The belief that nature allows neither genuine creation nor annihilation originated in antiquity; it is expressed in principles such as *nil posse creari de nihilo* and *causa aequat effectum*. In face of the experience of change, proponents of these ideas sought to uncover underlying constituents of reality that remained constant. The ultimate explanation of change, on this approach, is that change is only apparent. Among modern thinkers, Emil Meyerson is notable for advocating a kind of Parmenidean view on which change is illusory and "identity constitutes the essence of our understanding" ([1908] 1930, 402). Even when change is not altogether denied, it is generally believed to be constrained by some parallelism between earlier and later states, between cause and effect, between input and output. Descartes, who discovered (an early version of) the conservation of linear momentum, asserts:

> Now, it is manifest . . . that there must at least be as much [reality] in the efficient and total cause as in the effect of that cause. For where, I ask, could the effect get its reality from, if not from the cause? And how could the cause give it to the effect unless it possessed it? ([1641] *Meditations* III: 40; 1985 2: 28)[27]

Modern science has elaborated on these rudimentary intuitions about conservation in various ways. Classical mechanics led to the dis-

27. The translation (1985) is based on the original Latin text published in 1641; the brackets in this edition indicate insertions from the French version published three years later.

covery of the conservation of energy and linear and angular momentum, and QM has added further conservation laws. (Note that a theory can be indeterministic, like QM on the standard interpretation, and still impose strict causal constraints through its conservation laws.) In view of the fact that conservation laws are rooted in traditional ideas about causality, it is not surprising that the term *causality* has been used to refer to the applicability of conservation laws. For Niels Bohr, *causality* means the conservation of energy and momentum. In his oft-repeated claim that causal descriptions and spatiotemporal descriptions are complementary (that is, the accuracy of their joint application is restricted by Heisenberg's uncertainty relations), the term *causal description* should be understood in this way (and not, for instance, as connoting determinism). Explaining complementarity, Bohr states: "We have thus either space-time description or description where we can use the laws of conservation of energy and momentum. They are complementary to one another. We cannot use them both at the same time" ([1928] 1985, 6: 369).[28] Conservation laws and symmetries are inseparable members of the causal family. The causal function of conservation laws therefore also has bearing on the causal function of symmetry principles.

SYMMETRIES. Physicists place symmetry principles, which constrain the form of lower-level laws and guide theory construction, at the top of the hierarchy of physical laws. Symmetry considerations appear to be backed by a priori reasoning that resembles mathematics rather than physics. Their epistemic status is thus a matter of controversy. The connection with conservation laws, however, suggests that, to the extent that conservation laws are empirical laws that flesh out the causal structure of the world, so are symmetries. Although the connection between symmetries and conservation laws had been recognized earlier, it was proved by Emmy Nöther, who showed that, under a wide range of conditions, every continuous symmetry is correlated with a conserved quantity (Nöther 1918). In some cases, the

28. Quantum phenomena such as crossing a potential barrier seem to violate the conservation of energy and momentum. But ascribing definite energy and momentum values to the crossing particle would preclude its localization in space and time, hence we cannot "catch" it in the act of violation. This generates the complementarity Bohr invokes here.

connection between particular symmetries and other causal constraints is obvious. In the framework of STR, for instance, the principle of relativity, which is a symmetry principle, and the limit on the speed of signal transmission, which is a causal constraint, are closely linked. In other cases, gauge symmetries in particular, the connection is less obvious, and even debatable. I argue in chapter 5 that as a rule, symmetry principles function in the same manner as other causal constraints, and illustrate this claim by examining Pauli's exclusion principle. The connection between causation and symmetry is also conspicuous in Curie's principle, according to which symmetries manifested by a cause are inherited by its effects.

VARIATION PRINCIPLES. These principles single out the specific trajectories taken by physical systems. They determine, for instance, that light moves along the trajectory that takes the least time, that a particle follows the trajectory of least action, and that a freefalling body moves along a geodesic. Like symmetry principles, variation principles have a privileged status—they too are considered to be among the most general constraints on the form of theories. At first glance, variation principles appear to be teleological, and were indeed seen, when first discovered, as a demonstration of divine wisdom and benevolence. Over time, the teleological interpretation of these principles has given way to a causal understanding. Nevertheless, vestiges of the purposive impression seemed to linger. I will argue that, surprisingly, it was only in the context of quantum mechanics that the futility of the teleological interpretation could finally be established.

The foregoing list of causal constraints in physics introduces the constraints that will be examined in the coming chapters; it does not purport to be exhaustive. In addition, let me note two constraints that will not be thoroughly examined.

ASYMMETRY OF THE CAUSAL RELATION. Like determinism and locality, asymmetry is often considered an essential characteristic of the causal relation, and thus often referred to as "causality."[29] At the same

29. Frisch (2014) concentrates on this aspect of causation.

time, causal asymmetry has been contested on various grounds, especially its incompatibility with the fundamental laws of physics. The problem of whether and how causal asymmetry is related to temporal asymmetry is also much debated. Despite its centrality in the intuitive picture of causation, the asymmetry condition must be added "manually" in some of the leading accounts of causation, for example, the regularity and probabilistic accounts. On my pluralistic approach, the need to add this asymmetry to the other members of the causal family does not pose a problem. Moreover, causal asymmetry can be posited when focusing on individual processes and ignored when considering the general constraints imposed by conservation laws, symmetry principles, and variation principles. As they have no built-in asymmetry, these constraints play a causal role in controlling change, but not in controlling its direction. This tolerant strategy, I contend, is methodologically apt. Tolerance would be inappropriate, however, were the objection regarding the incompatibility between causal asymmetry and the fundamental laws of physics valid. But is it valid?

The incompatibility argument draws on the time-reversal symmetry of the fundamental laws of physics. A law is said to be time-reversal symmetric if whenever it allows a trajectory from event *c* to event *e*, it also allows the time-reversed trajectory from *e* to *c*. A common analogy is a film played backward: under time-reversal symmetry, we are unable to tell which film represents the actual course of events and which is the reversed film depicting a fictitious (though possible) course of events. The argument against causal asymmetry is that under the regime of time-reversal-symmetric laws, there is no observable difference between the two evolutions, and thus no reason to deem some events causes and others effects.[30] Consider a transition from an event *c* to an event *e*, and the following questions:

1. Do the fundamental laws allow us to retrodict the occurrence of *c* from the occurrence of *e* in the same way that they enable us to predict the occurrence of *e* from the occurrence of *c*?

30. When asymmetry is taken to be constitutive of causation, the argument targets the concept of causation tout court.

2. Do the fundamental laws allow the time-reversed transition from the occurrence of *e* to the occurrence of *c*?
3. Is there a sense in which *c* caused *e* but *e* did not cause *c*?

These questions are, in my view, distinct. Let us first consider questions (1) and (2). Laws that are deterministic and time-reversal symmetric yield an affirmative answer to both these questions, but this does not mean that the questions are equivalent. Had the laws been deterministic but not time-reversal symmetric, they would not necessarily sanction the reversed process, but could still allow retrodiction. On the other hand, under conditions of utter randomness, time-reversal symmetry could obtain despite the failure of prediction and retrodiction. As far as the incompatibility argument is concerned, however, the crucial point pertains to the relation between the first two questions and the third. From the affirmative answer to questions (1) and (2), the incompatibility argument concludes that question (3) must be answered in the negative. That is, it contends that if the time-reversed process can be predicted and is in fact allowed, there is no reason to take *c* to be the cause of *e* rather than take *e* to be the cause of *c*. But why should the laws' time-reversal symmetry exclude cause–effect asymmetry in the individual case? Over the last two weeks, I lost 3 pounds, but it would also have been possible for me to gain 3 pounds. (Indeed, given precise information about my diet and energy expenditure, these changes could have been predicted.) Does this mean that there is no fact of the matter as to what actually happened? Losing and gaining weight is a complex macroprocess involving much more than the fundamental laws of physics, but in principle, the point also applies to microprocesses. The fact that the laws of physics *allow* a process to unfold in opposed directions is compatible with the fact that, on any particular occasion, only one of these possibilities is realized. As it stands, therefore, the incompatibility argument, popular though it seems to be, does not refute causal asymmetry. More direct support for this asymmetry can be drawn from the discussion of Curie's principle in chapter 5.[31]

31. See Hemmo and Shenker (2012a) for an argument that anchors temporal asymmetry in the concept of velocity, and hence in fundamental physics.

MUTUALITY OF CAUSAL RELATIONS. Newton's third law states that if a body *a* exerts a force *F* on another body *b*, then *b* in turn exerts on *a* a force –*F* equal in size and opposite in direction to *F*. This law is violated in some physical theories (for instance, by the electromagnetic force), and is certainly not generally accepted outside physics. In philosophy, the idea of mutual action has sometimes been expressed more vaguely, requiring that if *a* can causally affect *b*, it must also be possible for *b* to causally affect *a*. Such stipulations appear, for example, in debating the mind-body problem. As I said, though, they are rarely encountered in science. A notable exception is Einstein's argument in support of the dynamic spacetime of the general theory of relativity (GTR). "It is contrary to the mode of thinking in science to conceive of a thing (the space-time continuum) which acts itself, but which cannot be acted upon" (Einstein 1922, 55–56). Here the mutuality constraint motivates the most revolutionary aspect of the new theory. According to Einstein, Newtonian mechanics violates our causal intuitions, for it allows space to act on matter, but does not countenance the reverse action, that is, the action of matter on space. GTR, according to which spacetime is shaped by the distribution of matter, while also determining this distribution, corrects this deficiency.[32] Einstein's use of the concept of causality in this context is somewhat idiosyncratic, but illustrates the fecundity of a concept of causality that is richer and more varied than the thin notion of causation debated in the philosophical literature.

This introduction has outlined the motivations for the book as a whole. Each chapter is largely self-standing, with the occasional slight overlap. Chapter 2 analyzes the determinism-stability relation as manifested in everyday contexts; chapter 3 analyzes it as manifested in physics. Both chapters show that the notions of determinism and stability are often conflated, giving rise to teleological thinking. Chapter 4 focuses on the

32. Note that this requirement of mutual causal influence does not involve the identification of individual cause-events and effect-events. As mentioned in note 13 above, DiSalle (1995) challenges Einstein's causal interpretation of the relation between matter and spacetime.

relation between determinism and locality, particularly in the context of quantum mechanics, where subtle payoff relations between these constraints are manifested. Chapter 5 examines symmetry principles and conservation laws. It illustrates how symmetry principles—despite their a priori appearance—function as causal constraints on a par with other members of the family of causal concepts. The "least action" principle is explored in chapter 6, which returns to the illusion of teleology, arguing that only within the QM framework is the principle's teleological appearance finally dispelled. Chapter 7 uses some of the results reached in earlier chapters to examine the relations between different levels of causality. It discusses reduction, emergence, and the intriguing possibility of lawless events in a deterministic world.

2

Determinism and Stability

THIS CHAPTER has two goals: first, it seeks to highlight the important role played by stability, a somewhat neglected causal notion, in our understanding of change. Second, it seeks to clarify the differences between stability and determinism, and the implications of these differences for understanding different types of change. As we will see, these differences are easily overlooked, especially, but not exclusively, in everyday discourse and "soft" sciences such as history. In this chapter, therefore, much of the discussion will be devoted to examples from outside the realm of physics, and in particular, to the case of historical explanation. In this context, the terms *necessity* and *contingency* are more common than the terms *stability* and *instability*, but as we proceed, it will become evident that there are close links between them. The role of stability in physical theories such as statistical mechanics is addressed in the next chapter.

The term *determinism* is relatively recent (its first OED citation is from 1846),[1] but the idea that events might be predetermined by the laws of nature, intrinsic *telos*, or divine will is ancient, and has been expressed in a variety of idioms, such as necessity, inevitability, and fate. The rubric of necessity, however, is notoriously ambiguous. I am not referring to necessity as used in the philosophy of logic and mathematics, where

1. The German *Determinismus* and the French *déterminisme* predate the English (OED) citation from Hamilton. Kant, for example, used the term several times in a moral-religious context in his 1793 *Religion innerhalb der Grenzen der bloßen Vernunft* (*Religion within the Limits of Reason Alone*).

it is fairly well delineated (even if its precise meaning and origin are still being debated), but rather, to the ambiguity of the rubric of necessity in ordinary discourse. Here, we often ask ourselves whether a certain event or chain of events was necessary or could have been avoided. What do we mean, for example, when we say that a defeat was inevitable? Do we mean that this particular defeat was causally determined by a very specific course of events (specific initial conditions) and the laws of nature? Or, alternatively, that a defeat more or less similar to the actual one would have taken place in any event, under a variety of different conditions? And further, does one of these interpretations correspond to the idea of the defeat's being *fated*? When Peirce defined truth as that which the community of investigators is *fated* to arrive at (Peirce 1878), did he have one of these senses in mind? We rarely bother to disambiguate the terms *necessary* or *inevitable* in such contexts, or to distinguish them from the notion of determinism, but I will argue that there are good reasons to be pedantic here, the most significant of which is the wish to explain away the appearance of teleology and directionality in various areas, including history and biology.

Before turning to analysis of the notion of necessity and its distinctive causal role, let me examine some of the characteristics of the more familiar notion of determinism. To begin with, we should note that traditionally, determinism finds its way into causal discourse through two distinct routes. First, on the most common and intuitive understanding of the causal relation, causes determine their effects in the sense that once the cause has occurred, the effect must follow. (There is, of course, always the caveat that this is only the case provided nothing interferes and prevents the effect from happening, but significant though this caveat is, it can be set aside for the moment.) What does "must" mean is this context? If we do not wish to think of the causal relation as a third entity over and above the related events, a little invisible chain that does the work, so to speak, then it seems natural to construe "must" as implying that recurrence of the cause guarantees recurrence of the effect. Hence the "same cause, same effect" principle, also known as the regularity principle, as constitutive of the concept of cause, and the close relation between causation and lawful behavior.

Undeniably, the term *same cause* harbors a problem, for we should actually speak of the same *type* of cause, and types (as Davidson pointed out and as I stressed in the previous chapter) involve descriptions. Thus if the causal relation is characterized via types, it too becomes description sensitive; this problem can wait as well.[2] The first route by which determinism enters causal discourse, then, ensues from the fact that lawful determination has often been taken to be a constitutive characteristic, or even a definition, or partial definition, of causation. Here the connection between causation and determination is so tight that we can understand why the terms *causation* (*causality*) and *determinism* are so often interchanged. Accepting this definition commits us to the view that in every case where there is a causal relation, the cause acts deterministically. Note that this principle is consistent with the existence of uncaused events. Determinism has also entered into the causal discourse by another route: the unconditional and much more general principle, also known as the universality principle, on which *every* event is determined by causes. This construal of determinism rules out the spontaneous occurrence of events, that is, chance. In this sense of determinism—the sense laypeople usually have in mind when using the term—determinism turns out to be identical to the notorious causal principle that led Russell and Norton (erroneously, in my view) to ban the concept of cause altogether.

The two principles, one of which asserts the regularity of causation, the other its universality, are quite distinct. On the one hand, it is possible that whenever the same (type of) cause recurs, the same (type of) effect recurs, but that uncaused, random events may also occur. John's being late always makes Jane angry, but she sometimes gets angry capriciously, for no reason whatsoever. On the other hand, it could be the case that every event has a cause, but the same (type of) cause does not invariably lead to the same (type of) effect. Jane never gets angry capriciously, but goings-on that occasionally make her angry—John's being late—do not always anger her. If, however, we *define* causation in terms

2. Davidson does not take singular causal statements to refer to types, and would deny that they are description sensitive.

of the regularity requirement, then a world obeying universality, that is, a world in which every event has a cause, is ipso facto a world in which the same (type of) cause leads to the same (type of) effect. The reverse does not follow; causation can be defined via the regularity requirement, but a world obeying regularity may still allow for chance—random events. Clearly, accepting the conditional statement that *if* there is a cause, it determines its effect(s), does not commit us to the general principle that *every* event is determined in this way. And further, while we may be free to define the causal relation in various ways, so that satisfaction of the regularity requirement is actually a matter of definition (we won't consider an event to be a cause if it doesn't invariably lead to the same effect), once we accept a particular definition of causation, the truth of the universality principle is no longer up to us.

Contemporary science does not define determinism in terms of the language of cause and effect. Instead, determinism is taken to require that two copies of a closed system that agree in their fundamental physical parameters at some time *t*, agree on the values of these parameters at all other times (or at least future times).[3] This definition captures both of the intuitions underlying the traditional characterization of determinism. Whereas the traditional requirements, universality and regularity, are distinct, the contemporary definition of determinism seeks to encompass both of them. If any two systems that are in the same physical state at one point in time continue to be in identical states at all other (future) times, deviation from either universality or regularity is excluded.[4]

Note that on the current definition, determinism is restricted to closed systems, making no provision for intervention by agents external to it or by the environment. This feature of determinism has the surpris-

3. The concept of a closed system is an idealization that cannot be fully actualized. Furthermore, the notion of the value of a parameter at a specific time also needs refinement, since some physical magnitudes, e.g., velocity, involve change over time, convergence to a limit, and so on.

4. The contemporary definition also excludes nonlocal influence on the system. Presumably, a nonlocal intervention could change some factors in the interior of the system without changing the boundary conditions, in which case the criteria for determinism would not be satisfied. The locality constraint is discussed in chapter 4.

ing result that determinism and causation may turn out to be *incompatible*. Recall the manipulation (interventionist) account mentioned in chapter 1, where causes are defined as factors whose manipulation changes the subsequent course of events. On this account, determinism and causation exclude each other: determinism only obtains in closed systems, whereas causation (in the manipulation/intervention sense) only obtains in open ones. Indeed, the argument regarding the incompatibility of determinism and causation has been made by Stachel (1969),[5] but given the broader conception of causation championed here, which does not define causation via intervention, this incompatibility need not concern us.

In many contexts, for example, that of the free will problem, the focus of philosophical interest is the truth of determinism "in reality," but it is more accurate to think of determinism (indeterminism) as a characteristic of *theories*. A theory is deterministic if it entails the above-stated condition, that is, entails that two systems that are in identical states at one point are identical throughout. Ascribing determinism (indeterminism) to the world is, then, simply shorthand for ascribing this property to our best theory of the world.

Determinism has sometimes been construed in epistemic terms: a theory is deterministic if knowledge of a system's initial conditions enables the prediction of any other state. This condition is stronger than the condition just discussed, since a state may be predetermined but nonetheless unpredictable due to difficulties in ascertaining the initial conditions, or due to complexity and computability considerations. In classical mechanics, the notorious three-body problem that led Poincaré (1905) to what we now call "chaos theory" involves a deterministic but unpredictable system. Hence determinism (as the term is used here) does not imply predictability.[6] It may also be useful to

5. Stachel does not refer to the interventionist account explicitly, but I believe that this is a fair characterization of his argument.

6. The connection between determinism and predictability is made by Laplace ([1814] 1994) in describing the omniscient demon, and is sometimes still assumed in the literature, e.g., in the *Encyclopedia Britannica* entry on determinism, updated in 2016. On the difference between the two concepts, see, e.g., Pitowsky (1996) and chapter 7 below.

introduce quantitative measures—degrees of correlation between earlier and later events—rather than make do with the binary division between determinism and indeterminism. The notion of indeterminism could then also apply to a wider range of situations, from very high, but less-than-perfect correlation, to genuine randomness—no correlation whatsoever.

These considerations make it clear that whether determinism is, in fact, the case depends on whether a deterministic theory of the world is true, and is thus an empirical question. Nonetheless, it has been argued that the truth of determinism is a conceptual issue that can be decided a priori. A simplistic version of this argument, but a version still worth rebutting, is the following (Taylor 1974). Assuming bivalence, every assertion is either true or false. Time is immaterial here. Hence just as it is determinately true or false (whether we know it or not) that at noon on 1.1.1900 it was snowing in Jerusalem, so it is determinately true or false (whether we know it or not), that at noon on 1.1. 2100 it will be snowing in Jerusalem. If it is true, the argument goes, it cannot be false (and vice versa), so whatever the case, it cannot be otherwise. This argument conflates logic and causation—categories that were not always properly distinguished by earlier philosophers, but which modern science keeps distinct, and I wish to keep distinct here. The truth of determinism hinges on the empirical question of whether the snowfall on a specified date is *causally* determined (determined by the laws of nature and the initial conditions, say); it has nothing to do with the "que será, será" tautology. Temporality is indeed immaterial to the validity of bivalence. The fact that it is true that it *was* snowing in Jerusalem at noon on 1.1.1900 does not entail that the snowfall that day was *causally determined*, and, similarly, the fact that it is true (in some nontemporal sense) that it *will* snow in Jerusalem at noon on 1.1.2100 does not render the snowfall that day causally determined.

We should note, parenthetically, that the same kind of logical error is behind the traditional apprehension that God's foreknowledge is incompatible with human freedom. Here, due to the theological setting, the fallacy is a bit harder to spot, but basically, it reflects the same conflation of logic and causation. If, as tradition has it, God is omniscient,

then the claim that God knows whether the sentence about snow falling in Jerusalem at noon on 1.1.2100 is true or false is simply a more picturesque way of saying that the assertion that it will snow in Jerusalem at noon on 1.1.2100 must have a determinate truth value. This does not entail that the snowfall is causally determined. Of course, if God's way of knowing is the human way, that is, calculating the result from the initial conditions and the laws of nature, then this knowledge presupposes determinism. But if God has noncausal ways of knowing (as Augustine, among others, contended), then God's knowledge of future events is as compatible with human freedom (in the libertarian sense of non-predetermination) as our knowledge of past events is.

With the crude argument for the a priori truth of determinism out of the way, let us look at Russell's more subtle argument to the effect that determinism can be trivially satisfied and therefore lacks empirical content (Russell 1913). Schematically, Russell's point is that all we need to satisfy determinism is a mathematical function that correlates the different states of a physical system over time. Given such a function, every state can be said to be determined by that function. But, Russell continues, such a function—indeed, more than one—will, as a matter of mathematical fact, always exist. For example, consider a single particle whose coordinates at time t are x, y, and z. Independently of how the particle moves, there will be functions $f_1(t), f_2(t), f_3(t)$ that correlate the earlier state x, y, z, with the later state x_t, y_t, z_t. The same is true for the universe in its entirety. Granted, the function for the whole universe would presumably be very complex, a function that we likely cannot specify, let alone calculate its values, but its very existence suffices to satisfy determinism. Russell admits that complex functions of this kind are not what science, as we know it, is after. Science seeks simple functions, but simplicity is not guaranteed by his argument. Although it establishes that, in principle, a deterministic theory exists, the argument does not establish the theory's usefulness, or set down a procedure for constructing it.

Understandably, Russell was not entirely happy with this conclusion. He therefore attempted to restrict the definition of determinism so as to make it a meaningful empirical property of theories, rather than a

vacuous one. The mathematical function that allows for determinism correlates the different states of the system *in time* (as a function of time). If we postulate the symmetry (equivalence) of temporal points, requiring the laws of nature to be independent of any specific time, the trivializing argument is thwarted, leaving the truth of determinism an open, empirical question. Russell was somewhat indecisive here but seems to have been inclined to take this route. If we go along with this inclination and admit only natural laws in which the time coordinate does not appear explicitly, we have a good example of symmetry considerations limiting the form of legitimate natural laws. From a different perspective, which Russell did not consider, such deliberation over the empirical content of determinism underscores Davidson's point regarding the crucial role played by classifying events into types by means of relevant and useful descriptions. The mathematical function that Russell takes to embody determinism correlates *individual* events, referred to by means of their spatial and temporal coordinates. As such it does not yield recurrent *types,* and is therefore useless as a tool of explanation and prediction—useless as a scientific law.[7] Moreover, the background of Russell's analysis was classical—he assumed without discussion that physical processes are continuous. In light of quantum mechanics, this assumption cannot be taken for granted. These considerations support the position espoused here; namely, that whether determinism is true is an empirical question. I would go further, and suggest that arguments purporting to show that determinism is a conceptual truth be viewed as a reductio ad absurdum of the definition of determinism on which they rest.

7. Maxwell notes that to make the "same cause, same effect" maxim intelligible, "we must define what we mean by the same causes and the same effects, since it is manifest that no event ever happens more than once, so that the causes and effects cannot be the same in *all* respects. What is really meant is that if the causes differ only as regard the absolute time and absolute place at which the event occurs, so likewise will the effects" (Maxwell [1877] 1920, 13). Maxwell makes it clear that only by postulating that time and place make no physical difference can we have types of events and lawful behavior. In "Funes, His Memory," Borges illustrates the importance of general types. Funes remembers events as instances, each in its unique singularity. At this level of detail, no two memories are alike, and nothing ever repeats itself. Funes's memories cannot be subsumed under general concepts and laws, which by their very nature require the omission of detail and distinctions (Borges 1998, 131–37).

Despite the empirical character of the question of determinism, it does occasionally happen that we have a *choice* as to whether to adopt a determinist theory, indicating that empirical considerations on their own do not always settle the issue. It can happen that there are empirically equivalent theories, only one of which is deterministic. David Bohm's deterministic quantum theory and standard (nonrelativistic) quantum mechanics (which is not deterministic) provide one example of this situation, and there are others, some of which arise in the context of gauge theories (discussed in chapter 5). In such cases there may be a trade-off between a theory's determinism and its other properties, in particular, its symmetry properties. As in other instances of choice between empirically equivalent theories, the argument for (or against) determinism in such cases depends on methodological considerations. Questions regarding the truth of determinism in specific physical theories have received considerable attention in the literature, especially from those who challenge the prevailing opinion that classical mechanics is deterministic whereas quantum mechanics is not. I will not address quantum mechanics in this chapter, and will uphold the traditional view that classical mechanics exemplifies a deterministic theory.[8]

Having expounded the notion of determinism, we can now turn to that of stability.[9] Some states are more stable than others: a ball inside a (not too large) spherical bowl is in stable equilibrium, whereas a ball atop the same bowl overturned is in a far less stable state. Stability is characterized in terms of the effect of small changes, a stable state being one to which a system returns after having been subjected to a small change. Upon being moved, the stable ball will always reach the bottom of the bowl, regardless of the specific trajectory of its movement inside the bowl, whereas the unstable ball, upon being dislodged from the top

8. The counterexamples to determinism in Newtonian mechanics, e.g., Earman's space invaders (Earman 1986) and Norton's dome (Norton 2007), arise from very specific initial conditions; see Malament (2008) for a critical examination of these counterexamples. Be that as it may, these exceptional cases have no significant bearing on the distinction between determinism and stability I am discussing here.

9. The remaining part of this chapter is a slightly revised version of my "Historical Necessity and Contingency" (2009) and is printed here with the kind permission of the publisher, Wiley-Blackwell.

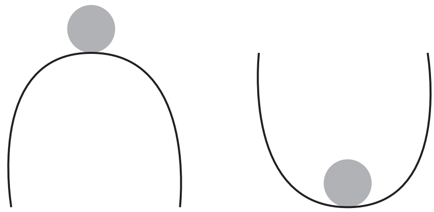


FIGURE 1. Stability and instability.

of the overturned bowl, will reach very different rest positions depending on the initial conditions of its motion. Yet both the stable and the unstable ball move in accordance with the same deterministic laws of Newtonian mechanics, indicating that determinism and stability are independent notions (figure 1).

Before further exploring the relation between these notions, consider some familiar events: a car accident, a meeting, a defeat. One question that can be asked about such events has to do with determinism. Was the event in question brought about by a deterministic process, or was it a random occurrence? Another question has to do with *alternatives* to the actual course of events. Could the accident have been prevented had the vehicle's speed been reduced? Would they have met had she not missed her flight? Would the battle have been lost had the weather been different, or had the commander managed to get some sleep that night? These questions are not about the events that actually transpired, but are, rather, about sets of *possible* events more or less similar to the events that took place. For obviously, the accident, the meeting, and the defeat that would have occurred had the initial conditions or intervening factors been different, would not have been the *same* events, but events of a similar—or more or less similar—kind. Even when we take the cause in a particular case to be a sine qua non condition (that is, the effect-event in question would not have occurred had the cause-event not taken place), it does not follow that, under different circumstances, an event similar to the actual effect-event would not have taken place. The

stormy weather was indeed part of the causal chain leading to the actual defeat, but if we consider the defeat to have been inevitable, we reckon that the battle would have been lost anyway, regardless of the weather. We must thus clearly distinguish questions regarding determinism—was the particular event in question predetermined?—from questions regarding stability and instability—would a small change have made a difference to the end result? Determinism, as we saw, means uniformity—recurrence of the *same* (type of) conditions ensures that a system evolves in the same way, but does not dictate that occurrence of *similar* conditions would result in the system's evolving in a similar way.[10] The independence of the question about similar conditions from that of recurrence under identical conditions is crucial for understanding the notions of stability and instability, both of which are compatible with determinism.

The literature on causation often suggests robustness or resilience as a desideratum. These terms, often used interchangeably, are ambiguous, however, chiefly because they are applied both to causal relations between individual events and to the laws instantiated by these relations. To avoid confusion, two senses of robustness (resilience) should be distinguished.[11] In one sense, robustness pertains to a law's scope. In another, robustness connotes a state or trajectory's stability (necessity, inevitability), that is, its resistance to perturbation. To illustrate the first sense, recall the problem of black body radiation that led Max Planck to his pathbreaking quantum hypothesis. On the one hand, Wien's displacement law was borne out by experiment for high radiation frequencies (high temperatures), but not for lower radiation frequencies. On the other hand, the Rayleigh-Jeans law was more accurate than Wien's for low radiation frequencies (temperatures), but, in diverging from empirical observations in the ultraviolet region, generated the "ultravio-

10. The independence of the two questions was stressed by Maxwell, who cites the maxim that "the same causes will always produce the same effects" and warns against confusing it with the maxim that "like causes produce like effects" ([1877] 1920, 13).

11. While not the only meanings of these terms, these are the meanings that matter in the context of characterizing causal relations and laws. The term *invariance* is also used to refer to the desideratum in question, as, e.g., in Woodward (2003, 15).

let catastrophe." By positing discrete quanta of energy, Planck's law escaped this divergence and achieved full agreement with experiment, thereby manifesting greater robustness than either of the earlier laws. Robustness in this sense of a law's having wide-ranging applicability is altogether distinct from robustness in the second sense, that of the stability of individual causal relations. As we have seen, robustness in the first sense, nomic robustness, is consistent with *in*stability: Newton's deterministic laws, which are robust in this sense, are consistent with the existence of unstable states. In neither of the two senses is robustness necessary for causation. Stability, in particular, is not a general characteristic of causal processes; it is not a general causal constraint.

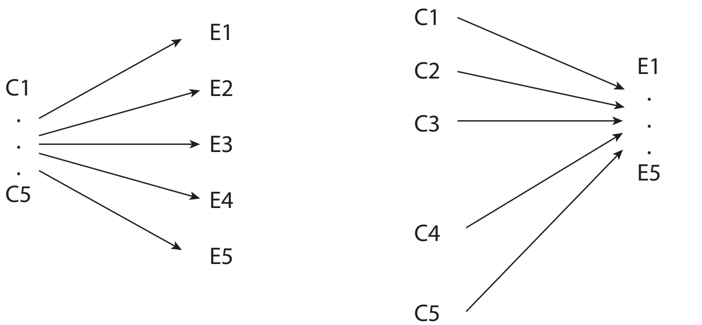
Historians and nonhistorians alike are often occupied with questions concerning the inevitability of certain events. (Historians are sometimes admonished to eschew counterfactuals and speculation as to scenarios contrary to the actual course of events, but rarely desist altogether.) We invoke such speculation, for instance, when distinguishing between causes and triggering incidents, implying that whereas the event in question would not have occurred had the cause been absent, the absence of the trigger could perhaps delay the effect, but would not ultimately avert it (or a similar one). When historians assert the necessity or inevitability of certain events, I take them to be making a claim much stronger than the mere statement that the events in question had a cause. Rather, the appeal to necessity implies that the effect-events were *overdetermined* by their causes, that they would have been (more or less) the same even had the cause-events been somewhat different.

Our opinions about alternatives to the actual course of events in history, as well as in quotidian discourse, are highly significant for decision making and the evaluation of actions. John is haunted by the thought that on the day Jane committed suicide, she was upset about his having canceled a planned visit. He would probably not blame himself as much for having canceled if he believed Jane's suicide to have been inevitable. The public/political sphere is equally replete with such assessments of historical contingency and inevitability. Take, for example, the different explanations put forward for the unequal status of

men and women in Western culture: some consider it a product of natural selection, a natural necessity, so to speak (note the juxtaposition of the notions of "nature" and "necessity" here); others consider it a mere social contingency. This divergence of opinion with regard to the etiology of the gender status gap in turn informs our views on specific gender-related sociopolitical issues. The more contingent we take the inequality to be, the greater our confidence in our ability to correct it, and (arguably) the greater our responsibility to try and do so.

In addition to this commonplace recourse to notions of necessity and contingency, the role of necessity or contingency in history is often systemically invoked by historians and philosophers of history. Hegel, Marx, and some of the Enlightenment *philosophes,* were, perhaps, the quintessential advocates of historical inevitability, while Nietzsche, Foucault, Berlin, and Rorty have vigorously championed contingency.

On the received view, necessity is associated with determinism and contingency with chance or randomness. By contrast, I suggest that contingency and necessity be understood in terms of stability, that is, sensitivity or insensitivity to initial conditions and intervening factors. This characterization takes contingency and necessity to be graduated, viz., it allows that they can be present to varying degrees. An event will be more contingent the more sensitive it is to initial conditions and intervening factors, and more necessary the less contingent it is, that is, the less sensitive it is to initial conditions and intervening factors. Thus, the defeat is more necessary if a similar defeat would have occurred in the absence of a number of conditions that in fact obtained—the storm, the sleepless night, the tactics chosen by the commander—and more contingent if changing these conditions would have changed the result significantly. As figure 2 illustrates, causes and effects, even when standing in a one-to-one relation, can still generate both necessity (when different causes lead to similar effects) and contingency (when similar causes lead to diverging effects). Admittedly, the notions of stability and instability have more precise application in physics than in historical contexts. In comparison with the balls in figure 1, our assessment of stability in the case of the accident or the defeat is highly speculative. My point, however, is conceptual; inevitability, like stability, is



Contingency: similar causes lead to different effects, manifesting high sensitivity to initial conditions.

Necessity: different causes lead to similar effects, manifesting low sensitivity to initial conditions.

FIGURE 2. Contingency versus necessity.

not synonymous with determinism. Hence, we should recognize two pairs of concepts, neither of which is reducible to the other: determinism versus chance, and necessity (or stability) versus contingency (or instability). Highly contingent processes can be perfectly deterministic. By the same token, processes interrupted by random events may still proceed to the final outcome without any change, for instance, when the system is approaching a stable equilibrium. Indeterminism is thus compatible with stability. In short, the concepts of necessity and contingency (stability and instability) are independent of the concept of determinism.

This analysis also affords a better understanding of the notion of *fate*, for fatalism seems appealing precisely when we have reason to believe that the "fated" event, such as Oedipus's murdering his father, would have occurred regardless of any intervening events, regardless of any action Oedipus might have taken. Fatalism is frequently identified with determinism, but if the aforementioned distinctions are respected, this is a category mistake. Determinism does not imply fatalism, just as it does not imply stability. The determinist is entitled to believe that *if* he carelessly raises his head, he may be hit by a bullet, whereas if he protects himself, he will be safe. The fatalist, on the other hand, typically

holds that a bullet will hit him (or that he will be safe) in any event, regardless of what he does. In taking steps to avoid the results of deterministic processes already in motion (but unknown to them), determinists are therefore consistent, whereas fatalists, inasmuch as they maintain that the fated result is too stable to be thwarted by their intervention, are right to deem such steps useless. When we invoke the rubric of necessity, we generally do so with regard to specific outcomes—the equilibrium, the accident, the defeat—that we take to be highly overdetermined. Necessity is thus associated with the image of an arrow pointing at such a specific event. This is not so in the case of determinism, where every outcome, stable or unstable, is taken to be predetermined. In the case of the rolling ball, we know which state is singled out as the equilibrium state. Literary works also explicitly or implicitly foreshadow the putative "fated" event. But in the absence of such information, how can the fatalist know whether any particular consequence is indeed fated? Whereas one can consistently hold, and it might even be true, that whatever happens is predetermined, it is virtually impossible that everything that happens, happens by "necessity" as we now understand this term. The fatalist thus faces the problem of singling out the fated event, a problem that need not concern the determinist.[12]

The conflation of chance and contingency, and of determinism and stability, is very common. E. H. Carr, for example, attacks what he calls "the crux of Cleopatra's nose" (Carr 1961, chap. 4)—namely, the idea that chance plays a significant role in history. He maintains, correctly in my view, that Cleopatra's nose, and other such oft-invoked intervening factors, provide no support whatsoever for the claim that history is a random string of events. Neither Antony's falling in love with Cleopatra, Carr argues, nor the results of the subsequent battle of Actium,

12. The *Stanford Encyclopedia of Philosophy* entry on causal determinism distinguishes between fate and determinism by linking fate to intention and determinism to natural causes. The distinction drawn here is more fundamental, as it renders them structurally different. The question of whether fate preempts human action was already discussed in antiquity, sometimes under the rubric "the idle argument." Notable examples are Aristotle (*Eudemian Ethics* II, 6 1222b 31) and Cicero (*On Fate*, 30). See Bobzien (1998) and Broadie (2007) for discussion.

were random occurrences. Indeed, on the analysis offered here, these events do not exemplify chance or indeterminism. Illustrating the significance of "small" interventions, they reflect causal connections characterized by a high degree of contingency. Similar examples can be found in Isaiah Berlin's seminal "Historical Inevitability." Seeking to create space for human freedom and responsibility, Berlin targets a whole cluster of philosophies of history, characterizing all of them as committed to historical inevitability. He conjectures that the appeal of inevitability is rooted in the desire to emulate the natural sciences so as to endow historiography with their elevated status. Berlin thus identifies inevitability with lawfulness and determinism. We have seen, however, that not every law-governed process manifests a high degree of necessity; in general, the laws found in scientific theories do not confer on the events they govern the kind of stability that justifies taking them to be inevitable. It seems to me that both Berlin's critique of historical inevitability and Carr's critique of chance would have been more effective had they distinguished between the various causal connections they group together.

Moreover, it is doubtful that the notion of human freedom that Berlin was eager to protect is of professional interest to historians. What they seek to discover, and what their evidence sheds light on, is not whether the historical agent acted freely and was thus morally responsible, but whether his or her actions *made a difference* to the course of history. When considering the assassination of Franz Ferdinand and his wife, Sophie, in June 1914 in Sarajevo, historians are interested in its impact on the ensuing outbreak of World War I, not the assassin's moral status. But making a difference to history, I suggest, is directly linked to the degree of necessity or contingency of the events in question. Human action is just one of several types of intervention in the course of history. Whereas highly contingent processes can be radically affected by such intervention, processes involving a high degree of necessity are far less susceptible to its impact. If, on our assessment, the war was inevitable, we will not ascribe the same weight to the assassination that we would if we believe the war was avoidable. By acknowledging

contingency, we create space for individual human beings to make a difference. It is the *historical impact* of individuals that historians defend or dispute, not their freedom of choice.[13]

Are assertions of necessity and contingency justifiable? It might seem that we can only come up with evidence related to the actual course of events, but not alternative scenarios. Basically, though, the epistemic problem we face when seeking to justify assertions about the latter is not qualitatively different from the problem we face vis-à-vis assertions about the former. The claim that World War I was inevitable could be corroborated by documentation of the relevant enmities, alliances, mobilization plans, and so on. And similarly for claims of contingency. Obviously, we are unable to quantify stability in history as we do in physics, but we can still reason about historical events, and about likely and unlikely alternatives to them, in an evidence-based manner.

The above distinctions are useful even when construed as applying to modes of *narrating* historical events, along the lines suggested by Hayden White, rather than to the historical events themselves. According to White (1973), historiography involves emplotment—organization of the historical material into narrative genres familiar to us from literature. Different configurations of the historical material may endow the same facts with different meanings. Surprisingly, White does not address the perspectives of necessity and contingency, but the literary patterns he identifies can be distinguished from each other by the degree of necessity they ascribe to the events in question. Tragedy is the paradigmatic manifestation of necessity. Writing a history of Napoleon as the story of a tragic hero entails conveying to the reader the sense that rather than being the master of his fate, Napoleon, like Oedipus,

13. Railton (1981) mentions the Sarajevo assassination in the same context, stressing the *modal* aspect of explanations that refer to a class of possible, rather than actual, events. Using the term *resilience* for what I call *necessity*, he also makes the connection between that concept and the concept of stability as used in science (251). I am grateful to one of the anonymous readers at Princeton University Press for bringing this paper to my attention. Given that the abovementioned confusions between determinism and stability are still quite common, it seems that Railton's insights on this issue (perhaps because they are peripheral to the subject of his paper) did not have the impact they deserved.

was in fact dominated by powers beyond his control. By contrast, writing the history of Napoleon from the perspective of a different genre, say the epic or the contemporary novel, will communicate a strong sense of contingency, leading to the conclusion that the historical outcome was shaped by even the smallest moves Napoleon made, that European history would have been entirely different had he not acted as he did. Greater complexity of presentation can be achieved by juxtaposing several genres, representing the perspectives of different players or observers. For example, the protagonist may manifest a highly contingent perspective while the narrator undermines that contingency by superimposing on the protagonist's perspective an overall tragic emplotment. The contrast between these perspectives then creates *irony*, the distance between the knowledge available to the protagonist and that available to the narrator or reader. In theological contexts, it is often the divine perspective that is invoked to expose the dimension of necessity behind the apparent contingencies of events; the story of Joseph and his brothers in the book of Genesis provides a classic example of a duality of this kind. Philosophers of history, too, are fond of the complexity and irony created by the juxtaposition of different perspectives, Hegel's cunning of reason being a case in point.

Returning to the notion of making a difference, recall that in addition to its intellectual and scientific aims, historiography has long been recognized as pregnant with critical insights. The critique of religion, morality, politics, and even science has benefited from historiographic research that has challenged the received myths and stereotypes associated with these institutions. This has made historiography a threatening discipline from the vantage point of those who wish to maintain the existing order or preserve the privileged status of a particular body of "truths," and a discipline pulsating with the promise of liberation for those who seek to engage in ideological housecleaning. Historiography as critique was advocated by Nietzsche, and practiced by thinkers as diverse as Spinoza, Marx, Feuerbach, and Foucault. And indeed, a critical effect is frequently generated by offering a change of perspective on the necessity or contingency of historical events, social structures, political arrangements, and so on. Wars, for example, are all too often

represented to the public as inevitable, the rhetoric of "necessity" serving obvious political ends. Reframing the events leading up to a war as highly contingent, suggesting that the war was escapable after all, will stimulate a more critical attitude to the war in question and perhaps other wars as well. Similarly, the bourgeois lifestyle is frequently lauded by those who embrace it as not merely morally commendable, but also *natural*, appropriate for human beings as such. The history of the emergence of this particular lifestyle may deflate this sense of naturalness, replacing it with the sobering perception that this particular way of life is but one among many.

In these examples, the change of perspective points in the direction of a greater degree of contingency, but changes of perspective in the opposite direction may also have sociopolitical implications. Sociobiology, for instance, strives to anchor our moral values and social practices in the evolutionary history of the species. Certain practices that are generally perceived as social constructs, and thus relatively contingent—habits of gift giving, say, or the greater sexual freedom granted to men than to women—emerge from sociobiological studies as more stable, from the evolutionary perspective, than their alternatives, and hence, despite appearances, bear the imprint of necessity. Occasionally, algorithms calculating "evolutionary stable strategies" (ESSs) are employed to buttress such arguments.[14] The abovementioned controversy over the gender status gap has involved the same struggle between the two kinds of perspective, with friends of necessity and sociobiology typically taking the more conservative stand and advocates of contingency leading the campaign for change. The perspective of contingency has become not only a lever for social critique and political change but also a manifesto for greater human independence in general. Further, the implications of contingency are increasingly appreciated in metaphysics and epistemology, and attempts, Kant's in particular, to identify the necessary elements of human thought and reason, increasingly eschewed. While Richard Rorty is widely considered the foremost

14. It should be noted, though, that even ESSs are not always stable in the above sense. A slight change can turn a winning strategy into a losing one.

proponent of the metaphysical plea for contingency, it is Michel Foucault who has contributed, more than any other historian, to dissolving the perspective of necessity vis-à-vis categories hitherto regarded as "natural," such as madness and sexuality. Foucault is quite explicit about the role of contingency in his vision; unlike critique in the Kantian sense of the term, the historical critique he is after, he tells us,

> will not deduce from the form of what we are what it is impossible for us to do and to know; but it will separate out, from the contingency that has made us what we are, the possibility of no longer being, doing, or thinking what we are, do, or think. (Foucault 1984, 46)

For Foucault, reframing the past as contingent is an ethical-political mission that enables us to forever change ourselves and forever change the world.

The notions of necessity and contingency as defined here are different from, but not unrelated to, their counterparts in metaphysics and modal logic. In the latter context, a proposition is considered necessary if it is true in all possible worlds—namely, true regardless of the features of a particular world, or simply, true no matter what; propositions that are not necessary are contingent. The historian has little use for this distinction—most truths she is interested in are clearly not true in all possible worlds, that is, they are contingent in the logical sense of the term. Relaxing the notions of necessity and contingency by construing them, not as binary, but as a matter of degree, allows us to make significant distinctions within the range of events that, from the logical point of view, are all indiscriminately classified as contingent. As we saw, it is useful to consider an event (more or less) necessary, not if it takes place *under all circumstances,* but if it is *relatively* insensitive to small changes in the circumstances under which it takes place.

Note that on the standard account of (logical and metaphysical) necessity, what one considers in contemplating the modal status of an event is whether *this very event* would take place in all, some, or no possible worlds. By contrast, I consider such ascriptions of modal status to implicitly refer to *sets of more or less similar events*. Recourse to notions such as kinship and similarity between events or trajectories is crucial,

for it implies description sensitivity. To assess degree of necessity, we need to know whether the same *type* of event would have occurred given a certain change or intervention. Our assessment therefore hinges on modes of sorting and individuation, on what we consider a type, or the same type. "The war was necessary" means "a similar kind of war would have occurred in any event," hence estimating the degree of necessity that should be ascribed to an event will depend on how broadly or narrowly we construe the type in question. A historian might believe a war would have started sooner or later, and consider the Sarajevo assassination a mere trigger. But if she uses more fine-grained individuation—a war in July 1914, a war triggered by an assassination, and so on—she will lower the level of inevitability ascribed to the war and reevaluate the significance of the Sarajevo assassination.

This observation develops the Davidsonian argument discussed in the previous chapter. Davidson distinguishes between causal and explanatory contexts: he considers the truth of singular causal statements independent of the description of the events in question, but contends that explanatory contexts, much like other intensional contexts, are description sensitive. This sensitivity is due to the fact that explanations contain laws that connect *types* of events rather than individual events. As the notions of necessity and contingency have to do with types, they too are description sensitive. In fact, given the vagueness of the notion of similarity involved in referring to sets of similar events, description sensitivity is even more conspicuous when assessing degrees of necessity and contingency than in Davidson's examples. Dependence on description is also a crucial, though often ignored, aspect of science, as will be shown in chapter 3.

It is a characteristic of stable states, we saw, that they can be reached from very different starting points. This feature creates a *structural similarity* between stability and teleology, and consequently, raises the danger of their conflation. A mechanical system approaching a stable equilibrium is undoubtedly very different from purposeful action. Nevertheless, there is a structural analogy between (at least one aspect of) goal-directed action and processes that lead to a stable equilibrium. Typical goal-directed behavior is flexible as to the means employed to

achieve a desired goal. I usually take the shortest route to campus, but if the road is jammed, I will choose an alternative route. Since rational beings can often achieve the same goal by various means, the outcomes of their purposeful actions manifest less sensitivity to intervening factors than do the outcomes of causal processes that are not self-adjusting in this way. In terms of structure, then, what our paradigmatic cases of necessity (the inevitable defeat, for example) and the actions of a rational being have in common is that the outcome is relatively impervious to minor changes in the initial conditions and intervening factors. Indeed, the very fact that flexibility and adjustability are familiar to us from the sphere of intentional action enhances our tendency to *project* intention onto cases where no conscious action is at play. The aforementioned fatalist, who takes a specific outcome to be inevitable, may be tempted to think of it as an end that must be achieved. From here it is just a small step to the personification of fate; not only is the outcome inevitable, but there is also a power that ensures that no matter how hard we try to elude our destiny, it will eventually catch up with us. Consider again the analogy at the root of such projections of intentionality: both features that epitomize necessity—indifference to any specific path, and salience of the endpoint—have parallels in goal-directed action. Note that even the schematic illustration of necessity, the many-one relation between different initial conditions and a single (type of) outcome, calls to mind an arrow or a direction. A mere structural analogy, then, leads us to associate substantially different phenomena: the structural similarity between purposeful action and natural processes that involve stable endpoints leads to the misguided injection of a direction, even a *telos*, into such processes.

Remnants of teleological thinking are still quite common in biology, inspired by phenomena such as *canalization*—the flexibility manifested by a living system in achieving the same kind of effect via different causal processes. Suppression of phenotypic variation, for instance, can be achieved by genetic canalization—insensitivity of a trait to mutations, and by environmental canalization—insensitivity of a trait to environmental variation. The structure of goal-directedness and canali-

zation that is characteristic of both human action and biological systems explains the temptation to think of an organism (cell, tissue, genome, species) as *working toward* a goal. The foregoing analysis of the patterns that resemble goal-directed paths may help us resist this temptation.

To appreciate the power of the "arrow of necessity" image, consider an example from evolutionary theory. Although Darwinism is generally acclaimed as providing a nonteleological account of evolution, questions regarding the significance of contingency in natural selection are far from settled. The Gould-Dennett debate reflects the emotionally charged disagreement over this issue.[15] The central thesis of Gould's *Wonderful Life* is that evolution (by natural selection and other factors, such as drift and catastrophes) is, to a considerable degree, a contingent process. Rewinding the tape of evolution, Gould maintains, would not reproduce the biosphere we have today.[16] Dennett takes natural selection to work more determinately, along the lines of an algorithm, and conceives Gould's contingency as the antithesis of the algorithmic process he posits. But as we saw, contingency versus determinacy is the wrong dichotomy; to counter Gould's affirmation of contingency, Dennett should have defended necessity, that is, the actual biosphere's stability and indifference to intervening factors. Had he done so explicitly, however, the difficulty of demonstrating his claim would have surfaced. Nevertheless, necessity and stability are implied in his arguments, and occasionally adduced directly. For example, he quotes the following passage from Darwin:

> More individuals are born than can possibly survive. A grain in the balance will determine which variety or species shall increase in number, and which shall decrease, or finally become extinct. (Dennett 1995, 41)

15. An entire chapter of Dennett (1995) is devoted to critiquing Gould's affirmation of contingency, ascribing to him a range of hidden anti-Darwinian agendas, from Marxism to religion.

16. Gould does not analyze the notion of contingency, but the metaphor speaks for itself.

Whereas Darwin's "grain in the balance" clearly suggests a high degree of contingency, in Dennett's immediately following paraphrase, the emphasis is on necessity:

> What Darwin saw was that if one merely supposed these few general conditions [variation, inheritance, limited resources] to apply at crunch time . . . the resulting process would *necessarily* lead in the direction of individuals in future generations who tended to be better equipped to deal with the problems of resource limitation that had been faced by the individuals of their parents' generation. (Dennett 1995, emphasis in original)

Similarly, Dennett quotes a review of Gould by Maynard Smith:

> In Gould's "replay from the Cambrian" experiment, I would predict that many animals would evolve eyes, because eyes have in fact evolved many times, in many kinds of animal. I would bet that some would evolve powered flight, because flight has evolved four times, in two different phyla; but I would not be certain, because animals might never get out on the land. But I agree with Gould that one could not predict which phyla would survive and inherit the earth. (Dennett 1995, 306)[17]

Dennett reconstructs this passage as follows:

> Maynard Smith's last point is a sly one: if convergent evolution reigns, it doesn't make any difference *which phyla* inherit the earth, because of bait-and-switch! Combining bait-and-switch with convergent evolution, we get the orthodox conclusion that *whichever* lineage happens to survive will gravitate towards the Good Moves in Design Space, and the result will be hard to tell from the winner that would have been there if some different lineage had carried on. (Dennett 1995, 306, emphasis in original)

This reconstruction, especially the image of evolution gravitating toward optimal solutions, assumes far more necessity than Dennett can

17. The quote is from Maynard Smith (1992), a review of Gould's *Wonderful Life*.

actually demonstrate. He thinks the "good moves in design space," such as eyes, will appear again and again. In a sense, then, the eye's evolution can be seen as inevitable. But there are many different kinds of eyes; do they count as the *same* move? The degree of necessity, as we saw, can be exaggerated or downplayed, depending on the description one chooses. Dennett is aware of the fact that description matters, as evidenced by these remarks:

> *Our* appearance? What does that mean? There is a sliding scale on which Gould neglects to locate his claim about rewinding the tape. If by "us" he meant something very particular—Steve Gould and Dan Dennett, let's say—then we wouldn't need the hypothesis of mass extinction to persuade us how lucky *we* are to be alive. . . . If, at the other extreme, by "us" Gould meant something very general, such as "air-breathing, land-inhabiting vertebrates," he would probably be wrong. . . . So we may well suppose he meant something intermediate, such as "intelligent, language-using, technology-inventing, culture-creating beings." This is an interesting hypothesis. . . . But *Wonderful Life* offers no evidence in its favor. (Dennett 1995, 307, emphasis in original)

Gould is quite specific: by "us" he means *Homo sapiens*, and the reference class is other branches of the human lineage, such as *Homo erectus*. Dennett thinks Gould's hypothesis is wrong, but his evidence is just as inconclusive as Gould's. It must be acknowledged, it appears, that at this point we know very little about the precise degrees of evolution's necessities and contingencies. The conflation of chance and contingency, however, adds confusion to ignorance. Dennett understands Gould's explanation of evolution as based on randomness, to which he responds with a rhetorical question: "Was it truly *just* a lottery that fixed *all* their fates?" (Dennett 1995, 304, emphasis in original). Gould does (inadvertently) mention Lady Luck, but his argument clearly centers on contingency rather than chance. The pitfalls in this exchange are those pointed out above; they could have been avoided by use of more accurate terminology, terminology that preserves both the distinction between stability and determinism and that between instability and chance.

In this chapter I have presented stability as a distinct causal category, independent of the category of determinism with which it is often conflated. I have identified a structure typical of processes that involve stable endpoints, whether or not they are goal directed. This structure, I argued, has been a source of confusion about directionality, teleology, and even fate. The next chapter illustrates how tricky the move from stability to directionality can be in the physical sciences.

3

Determinism and Stability in Physics

CHAPTER 2 ELUCIDATED the notion of stability as a distinct member of the causal family, to be distinguished, in particular, from determinism. Having examined the causal role of stability in daily discourse, we can now turn to the physical sciences, where the notions of stability and instability, no longer camouflaged in the language of necessity and contingency, are widely used in a variety of contexts, from chaos theory to quantum mechanics. For the physicist, questions about the stability of states, orbits, and structures are no less fundamental than questions about determinism. Understanding the structure of matter, whether at the smallest scale—elementary particles—or the largest—stars and galaxies—involves understanding the stability of certain structures and the instability of others. Quantum mechanics, for example, had to face the question of the stability of atoms, a rather ad hoc answer to which was given by Bohr's 1913 model of the atom. Initially, the model was based on an analogy with the solar system, with electromagnetic forces taking the place of gravity, and electrons conceived as orbiting the nucleus like planets. Bohr soon realized, however, that according to (classical) electromagnetic theory, electrons orbiting the nucleus would radiate energy, and as a result, their orbits would decrease continuously until they hit the nucleus. On this classical picture, there could be no stable atoms. Bohr therefore came up with a quantum model, according to which only discrete orbits—stationary states—are allowed, and

further proposed that electrons in these stationary orbits do not lose energy through radiation. The model was remarkably successful in accounting for the observed spectra of atoms, but its underlying quantum conjecture, which explained why atoms do not collapse, had no independent theoretical basis. Provision of such a basis, that is, accounting for the atom's extraordinary stability, was one of the chief goals of quantum theory in the 1920s.

Questions regarding stability also arise at the other end of the mass scale. Newton was concerned about the stability of the solar system, a concern that gave rise to the celebrated three-body problem and to Poincaré's development of what we now call chaos theory. In the 1930s, Chandrasekhar pondered the stability of stars, wondering why they didn't collapse under the inward force of their own gravity.[1] Perhaps, he speculated, they did, turning into what Wheeler later called "black holes." Stability is also a central concept in other areas of physics, such as thermodynamics and hydrodynamics. Indeed, stability is integral to our understanding of change on every level of the physical world, and is thus an irreducible member of the causal family.

The stability of a state or dynamical orbit is characterized by its response to small perturbations—the smaller the effect of perturbation, the more stable the state or the orbit. The mathematical theory of stability introduces finer distinctions, such as the following: small perturbations of a stable orbit will result in "nearby" orbits (where the distance between orbits can be mathematically defined and quantified); small perturbations of an asymptotically stable orbit will result in orbits converging on the original orbit. Perturbation can also result in orbits that do not converge at all or that are repelled from the original orbit or converge on a distant one. Various combinations of these possibilities (for different kinds and directions of perturbation) also occur. The details of the mathematical theory of stability, which grounds the theory of chaotic systems, need not concern us here.[2] For the purpose of pro-

1. The problem had also been raised independently by Wilhelm Anderson and Edmund Clifton Stoner. More on Chandrasekhar's work and its relation to Pauli's principle can be found in chapter 5.

2. The mathematical theory compares different, possibly infinitely close, states and orbits. No *actual* perturbation is at issue, hence there are no temporal implications. Physics studies

viding an exposition of the causal spectrum, what is important is the claim that determinism and stability are independent causal notions, a claim I made in the previous chapter and elaborate on in this one. The deterministic laws of classical mechanics are compatible with both stable and unstable states and trajectories. Similarly, as we will see in chapter 6, an indeterministic theory such as quantum mechanics may also allow both stable and unstable states and trajectories.

We saw that the ambiguity of the notion of necessity in nonscientific contexts blurs the important distinction between issues pertaining to determinism and lawfulness and issues pertaining to stability and resistance to small changes. We also saw that stability and directionality can be readily conflated, lulling us into teleological thinking. This confusion will continue to engage us. Here I focus on statistical mechanics, a theory that pivots on questions of stability and directionality, and therefore serves as a good case study. Indeed, accounting for the observed directionality of numerous physical processes has been one of the goals of statistical mechanics since its inception at the end of the nineteenth century, a goal that, arguably, has yet to be fully achieved. The example of statistical mechanics adds a new dimension to the discussion of stability, as it involves complex relations between different physical levels: the fundamental level of microprocesses, and the thermodynamic level supervening on these processes. It thus links the question of directionality to questions about the relation between different causal and explanatory levels, laying the groundwork for the discussion of reduction and emergence in chapter 7. In devoting a chapter to statistical mechanics, I do not presume to contribute to the ongoing debate about the foundations of this theory, but rather to use this example to extract and refine the conceptual relations between the causal notions of stability and determinism, and their putative contribution to the explanation of directionality.

In everyday experience, we regularly encounter instances of cooling off, spreading, and mixing, that is, instances of systems evolving in time

---

actual perturbations of states and orbits, and their likelihood. It therefore applies the mathematical theory to the effects of real perturbations, taking into account the perturbations' frequency, which directly affects stability over time intervals.

from unstable states that are short-lived and easily destroyed to more stable equilibria. These processes appear to be completely irreversible—we never see them spontaneously "rewinding." The latte that I left on my desk cooled to room temperature, but will never spontaneously draw heat from the environment so as to again be steaming hot, nor will it spontaneously separate into coffee and milk. Although classical mechanics could offer no explanation of irreversibility, explanatory progress was made in the framework of classical thermodynamics. Thermodynamics explains irreversibility in terms of the second law of thermodynamics, which asserts that when a system is in equilibrium, or interacts with the environment only adiabatically, its entropy does not decrease.[3] The definition of entropy, a physical magnitude that, under these idealized conditions, can change in one direction only, and can thus account for thermodynamic irreversibility, was a great innovation. Furthermore, in classical thermodynamics, the *stability* of the stationary state of maximal entropy and the *direction* of entropy increase were closely connected. Remarkably, though, more than a century of research has not yet produced a fully satisfactory explanation of the stability–directionality nexus. In a nutshell, the problems impeding the emergence of such an explanation pertain to the connection between thermodynamics and mechanics. The more feasible it became to unify thermodynamics and mechanics within the more comprehensive framework of statistical mechanics, the less feasible it became to uphold the classical explanation of irreversibility in terms of the second law. Let me elaborate.

We have already seen that stability is typically associated with a many-one relation, as in the case where various types of initial conditions lead to the same type of final state. Consider the crucial role of many-one relations in statistical mechanics. In principle, we can have different descriptions of a thermodynamic system. In particular, we can conceive of a microdescription specifying the values of every physical parameter of each of its constituent particles, as well as a macrodescription in terms of its macroobservables, such as its pressure, volume, and

3. In an adiabatic process, there is no heat exchange with the environment.

temperature. As it happens, the former description is unavailable to us, macrocreatures that we are, but the latter is readily obtained. Classical thermodynamics was formulated in terms of macrodescriptions alone, whereas statistical mechanics seeks to connect the two levels of description. The recognition that such a connection exists was driven by the kinetic theory of heat, on which heat is an expression of the incessant movement of huge numbers of particles that move and interact in accordance with the laws of classical mechanics. For some macroscopic parameters, the connection to microproperties is relatively clear—the correlation between the pressure exerted by a gas on its container and the average impact (per unit area) of microparticles on the container, for instance, is quite intuitive. But for other macroproperties, entropy in particular, the connection is more tenuous. Recovering the notion of entropy from the physics of the microlevel is essential for recovery of the second law of thermodynamics, and constitutes a major challenge for statistical mechanics.

The fundamental insight underlying the connection between entropy and the microlevel was that macrostates are multiply realizable by microstates. The implication is that in general, a detailed description of a system's actual microstate plays no role in the thermodynamic description of its macrostate, for the same macrostate could have been realized by numerous different microstates. What is crucial for the characterization of a macrostate in terms of its microstructure, however, is the *number of ways* (or its measure-theoretic analogue for continuous variables) in which a macrostate can be realized. As long as we have access to these numbers (or their measure-theoretic analogues) and can use them to distinguish between different macrostates, the fact that the detailed description of the actual microstate remains hidden is no obstacle. The standard formalism that captures this relation between microstates and macrostates is the representation of the former by points, and the latter by regions, in the 6N-dimensional phase space (where a point represents a microstate of the entire system in terms of 6 coordinates for each of its N constituent particles; in general, 3 coordinates for position and 3 for momentum). Each macrostate is realizable by all the microstates corresponding to points that belong to the volume representing

this macrostate; clearly, this is a many-one relation that can vary enormously from one macrostate to another.[4] This insight led to the identification of the volume representing a macrostate in phase space with that macrostate's probability, and to the definition of entropy in terms of this probability.[5] On this conception, the maximal entropy of the equilibrium state in thermodynamic terms is a manifestation of its high probability in statistical-mechanical terms. These definitions constituted a crucial step in transforming thermodynamics into statistical mechanics, that is, in recasting a deterministic theory of the behavior of macrosystems, formulated in terms of thermodynamic macroproperties, as a probabilistic theory of the behavior of multiparticle systems, formulated in terms of their underlying mechanical properties. A crucial step, but certainly not the end of the story.

Before mentioning some of the problems engendered by this transformation, we should note the features common to the analysis of stability in everyday discourse (discussed in chapter 2) and the probabilistic analysis of a thermodynamic system. As we saw, one such commonality is a many-one function correlating multiple arguments with a single value, say, the many routes that would have led to a defeat with that outcome, and sets of possible microstates with the macrostate they realize. Admittedly, in the former context we have in mind only a handful of possibilities, whereas in the latter we are considering statistical mechanics' vast aggregates of particles. But in both cases, the many-one relation is the formal manifestation of the *insensitivity* of the function's range to many of the features that distinguish its arguments from each other. As outlined above, the key feature of a macrostate—its probability—is relatively indifferent to many of the details of the microstates compatible with this macrostate. A second commonality is

4. As the number of points is infinite, technically, the "size" of a macrostate should be formulated in terms of its measure rather than in terms of a number. Present-day writers emphasize that using the Lebesgue measure for probability in this context is not the only possibility, hence in doing so, a nontrivial, albeit intuitive, assumption is being made. See, e.g., Albert (2000), Pitowsky (2012), Hemmo and Shenker (2012b).

5. This ahistorical reconstruction of the rationale that led to the probabilistic definition of entropy is closer to Boltzmann's conception than to Gibbs's. For the moment, I am deliberately overlooking the differences between their approaches; their relevance to the problem of the meaning of probability is discussed in note 7 of this chapter.

description sensitivity; macrostates do not descend from heaven with fixed identities, but acquire an identity from the description we use in referring to them. This does not mean macrostates are fictions—they are as real as microstates—but it means that if we are to understand their behavior *as macrostates,* that is, to discover the laws governing their behavior as macrostates, they must be identified in a useful way, and this identification cannot be read off their microproperties alone. (The significance of this point will be discussed below.) Third, we have seen that many-one relations can be misleading, and may invite the projection of directionality.[6] In statistical mechanics, in contrast to history and biology, the problem has not been overt teleology, but rather the temptation to mistake the probabilistic characterization of macrostates for a full-blown explanation of a thermodynamic system's *evolution* toward certain macrostates. For it seems natural to assume that the system will evolve (or most probably evolve) from less probable to more probable states, rather than vice versa, and that *nothing more* is needed to account for thermodynamic behavior. But how is this seemingly reasonable assumption to be justified?

To see what is involved in such a justification, consider three interconnected problems that statistical mechanics has wrestled with from early on.

1. The meaning of probability in statistical mechanics.
2. The connection between probability and a system's dynamics.
3. The origin of directionality.

1. The Meaning of Probability

Defining the probability of a macrostate in terms of the number of its possible realizations (or that number's measure-theoretic analogue) raises the question of the meaning of probability in this context, and

6. Conceptually, the many-one relation is not a necessary condition for directionality. In principle, one can imagine individual trajectories that have a built-in direction. In fact, the least action principle was initially interpreted as imposing such a direction on the trajectory of a mechanical system. But the association between directionality and the many-one relation is general enough to merit consideration, and can (as will be shown in chapter 6) replace built-in directionality even in the case of the least action principle.

the relevance of this probability to the actual evolution of an actual system. A plausible account of probability, it would seem, would involve an ensemble (possibly infinite) of similar systems where the probability that an individual system has a certain property is determined by the size of the fraction of the ensemble whose members have that property.[7] On this understanding of probability, in such an ensemble of systems we would expect to find more of the ensemble's systems in probable macrostates than in improbable ones. But how is this expectation reflected in the case of an *actual* system? Given an actual system that happens to be in an *im*probable macrostate, it would certainly not follow from the ensemble model that such a system can be expected, in time, to move from its improbable state to a more probable one. If our goal is to account for the evolution of an individual system (in particular, its evolution toward equilibrium), there is a gap in the argument.

The concern, to put it slightly differently, is that in making the assumption about evolution toward more probable states, we are conflating different kinds of possibility. The alternative ways of realizing a macrostate correspond to the possibilities of many different systems, that is, systems differing in their microstates. But these are not possibilities open to any individual system, which presumably has only one possible trajectory, the trajectory dictated by the deterministic laws of classical mechanics and the system's initial (micro)conditions. What difference do other possibilities make to any individual system if none of them are open to it? That these possibilities are possibilities of a different sort should not, in itself, disturb us, for we can decide which sense

7. This ensemble interpretation of probability is the basis for Gibbs's approach to statistical mechanics. Whereas Boltzmann's approach is more focused on the dynamical evolution of the individual system, in Gibbs's statistical mechanics it is the mathematical precision of the notion of probability, and thus the ensemble of identical systems, that takes priority. The problem I address is that of integrating the temporal evolution of thermodynamic systems with the Gibbsian approach. Boltzmann's approach, on the other hand, confronts the parallel problem of tying the evolution of an individual system to a reasonable account of probability. A significant difference between the two approaches is that on Boltzmann's approach, macrostates supervene on microstates, that is, although macrostates are multiply realized by microstates, each microstate belongs to a *single* macrostate. On Gibbs's ensemble approach this is not generally the case.

of "possibility" is appropriate in a particular context, but insofar as we are seeking a connection between the two descriptions, the description of the ensemble and the description of an individual system, the ambiguity is worrisome.

The problem as such is not peculiar to statistical mechanics; frequency and ensemble interpretations of probability are notoriously forced when applied to an individual case. What makes the problem more vexing in the context of statistical mechanics is that as physicists, we are definitely interested not only in probabilistic information about the distribution of systems (in a given ensemble) over different macrostates, but also in the evolution of individual systems and the mechanisms governing this evolution. Moreover, the irreversible evolution of such individual systems is precisely what classical thermodynamics was successful in explaining. The point of its transformation into statistical mechanics would be jeopardized were statistical mechanics to fail where classical thermodynamics had succeeded.

Consider a deck of cards. Its "microstates" are all the possible arrangements of the cards in the deck. Its "macrostates" are sets of such microstates that have a certain "macro"-property (that is, sets of microstates under a certain description), such as arrangements with all the kings on top; arrangements where red and black cards alternate; ordered, or disordered, decks; and so on. As in statistical mechanics (though in comparison, the number of card arrangements is miniscule), these "macrostates" vary in the number of microstates that realize them. Relatively few arrangements realize an ordered "macrostate," and the vast majority belong to the disordered "macrostate." Invoking the aforementioned expectation, if we consider all possible arrangements, we can expect to find the majority of decks that we encounter in the disordered "macrostate." But if the deck of cards was just purchased, this information is irrelevant; in particular, the said expectation does not imply that an ordered deck of cards left in a drawer will spontaneously evolve into a disordered one. Of course, the probabilistic reasoning becomes highly relevant when we *shuffle* the deck, thereby moving it through a series of different microstates. It tells us that in doing so we will most probably end up with a disordered deck—the

most probable "macrostate"—and that it is extremely unlikely that by shuffling a disordered pack we will return to the improbable ordered "macrostate." Note, however, that shuffling involves external intervention, not the *spontaneous* evolution of the system that statistical mechanics set out to explain. Nonetheless, if shuffling enables us to close the gap between the different kinds of possibility, an analogous process could, perhaps, play the same role in the context of statistical mechanics. But the question of whether such an analogous process exists cannot be answered by probabilistic considerations alone, and entails anchoring the probabilistic considerations in the system's underlying dynamics, which leads to the second problem.[8]

2. Probability and a System's Dynamics

The need to link the relative stability of a system's particular state with the system's underlying dynamics and boundary conditions doesn't arise only in statistical mechanics. The ball's stability in the bowl, and instability atop it, for instance, are also anchored in a specific dynamic—in this case, Newton's laws of motion and gravitation. Unless such dynamical considerations are taken into account, there is no explanation of why one position is more stable than another. And yet, as already emphasized, stability as such is compatible with different kinds of dynamics, and conversely (depending on various conditions), the same dynamics can underlie stable as well as unstable configurations. Hence establishing the connection between any particular manifestation of stability and the specific dynamics of the system at hand is of critical importance. But recall the significance of description: before one can establish such a connection, it must be decided what counts as the *same* state (the state the system stays in, or returns to), and thus one must begin with the classification and individuation of states. Once such a classification scheme is in place, we can evaluate the probability of dif-

8. Historically, the probabilistic derivation of the Maxwell distribution of velocities (by Maxwell and then by Boltzmann) was tied to dynamic considerations from day one, but to facilitate presentation of the conceptual problems in a nontechnical way, I am not following the historical development.

ferent states and their degree of stability. The system's dynamics and the state descriptions play different roles: the underlying dynamics explains how the states we are positing can be realized; the descriptive categories, in turn, define the "size" of these states and thus potentially have bearing on their probability. We don't always go into the details of the system's dynamics, but we must have at least a schematic picture. When we shuffle our deck of cards, the shuffling (whose details we can ignore) provides the dynamics, and the sole assumption we make about the shuffling is that it does not favor any particular arrangement. The role of the categories is to assemble the "microstates" into "macrostates"—an ordered or unordered deck of cards, and so on. These categories define the size of the "macrostate" we have singled out for inspection, and affect the probability of getting it, in the long run, via shuffling.

There is a property—*ergodicity*—that could solve both problems I have mentioned: relating probability to a system's temporal evolution, and supplying the dynamic underpinnings of this evolution. A system is ergodic if it passes through every microstate consistent with its mechanical constraints, say, its energy (its Hamiltonian) and the space it is confined in, say, a container. These constraints can be satisfied in numerous ways,[9] and in general pick out a "hypersurface" within the phase space. If ergodicity could be derived from the laws of Newtonian mechanics, it would paint a picture on which the trajectory of the system obeying these laws wanders the surface picked out by the constraints in such a way that, in the long run, it *actually* visits all the points it can *possibly* visit. If ergodicity were to obtain, it would constitute another application of the principle that Cox and Forshaw (2011) propose as a basis for Feynman's conception of quantum mechanics: anything that can happen, does happen—see chapter 6. It would now make sense to look at the probability of a macrostate as reflecting not only the fraction of systems in a *hypothetical* ensemble found in that macrostate, but also the actual time spent in that macrostate by each individual system: a system would spend more time in more probable macrostates. Ergodicity would therefore link the probabilistic analysis of entropy to the

9. E.g., the same total energy is compatible with different distributions of velocities among individual particles.

temporal evolution of individual systems, and anchor this linkage in the underlying dynamics. As a matter of mathematical fact, however, the conditions for ergodicity cannot be satisfied in systems of dimension greater than 1. There is a family of weaker conditions that have been shown to be mathematically possible and to support, to various degrees, the "wandering" or "stirring" picture of the system's visiting numerous possible microstates. The first example of such a weaker notion is due to Birkhoff and von Neumann, who showed that under certain conditions the trajectory is "dense," so that although it does not pass through every point, it does get arbitrarily close (in the technical sense of this term) to every point, and further showed that, while this property does not hold for all trajectories, it holds for most of them (again in the technical sense of missing at most a measure zero set).[10] The question of whether there are any concrete physical systems that meet (or approximately meet) the conditions for ergodicity in the Birkhoff–von Neumann sense, or the conditions required for any of the weaker properties in the "ergodic hierarchy," remains unresolved. It divides scholars into two camps: those who remain confident in the physical reality of some descendant of ergodicity that supports the "stirring" picture, on which the system's dynamics drive it toward more probable macrostates (thereby establishing a relation between possible and actual evolutions of a system), and those who do not.[11]

10. See, for example, Birkhoff (1931) and von Neumann (1932). For a detailed discussion of the historical questions of whether Boltzmann implicitly or explicitly assumed ergodicity, or a weaker condition such as quasi-ergodicity, and at which specific points in his argument such assumptions might be relevant, see Uffink (2007) and the literature cited there. On the ergodic hierarchy of properties such as "mixing" and the Bernoulli property, see Sklar (1993) and Uffink (2007).

11. Albert (2000) criticizes ergodicity as not only not even approximately true, but also *irrelevant* to statistical mechanics. He argues that the context of the ergodic approach is epistemic, which renders it subjective and as such unsuitable to be a physical explanation. The difference between my notion of description sensitivity and subjectivity will be explained momentarily. Uffink (2007) points out that there are properties in the ergodic hierarchy, such as mixing, that make various assumptions about initial conditions—in particular, the assumption of their typicality—more reasonable. He therefore contends that Albert's approach to statistical mechanics, which invokes the "past hypothesis," would benefit from adducing quasi-ergodic considerations.

For our purposes, two points should be noted. First, in general, the techniques used in deriving such ergodic properties require various methods of partitioning the phase space into "cells" (a procedure referred to as "coarse graining") whose microstates are indistinguishable from one another from the perspective of that partition. Second—and this feature further underscores the significance of the description we choose—a description of a system's evolution in terms of the behavior of these cells may differ considerably from a description of its evolution in terms of its microstates and underlying dynamics. In particular, stochastic behavior can supervene on an underlying deterministic dynamic.[12]

It has been claimed that the probabilities of statistical mechanics are epistemic or subjective. By contrast, I have stressed description sensitivity. To see the difference between these claims, let's take another look at the definition of entropy in statistical mechanics. Statistical mechanics, we have seen, establishes a connection between different levels of description—higher-level description in terms of macrostates, and basic description in terms of microstates. This connection grounds the familiar claim that thermodynamics is being reduced to mechanics. In the ideal case of reduction, the entities, properties, and laws of the higher level are correlated with entities, properties, and laws expressed entirely in the vocabulary of the basic level. The aforementioned definition of the pressure exerted by a gas on its container, a definition framed in terms of the impact of individual particles, is a typical example. The definition of entropy, however, deviates from this ideal. Rather than being defined exclusively in the microlevel vocabulary, entropy is defined as a *relation* between the two levels—the number of ways in which a higher-level state can be realized (or its measure-theoretic analogue in terms of phase space volume). Evidently, there is no physical property of an individual microstate, nothing identifiable by looking at the microstate *as a microstate*, that can disclose the property of its belonging to a particular macrostate. And there is nothing that distinguishes the

12. Uffink emphasizes this achievement: "One of the most important achievements of ergodic theory is that it has made clear that strict determinism on the microscopic level is not incompatible with random behavior on a macroscopic level" (2007, 1017).

set of microstates that realizes a particular macrostate other than the very fact that its members realize that macrostate. Hypothetical micro-creatures (conscious molecules, say) who perceive only microproperties would not be able to distinguish that set from other sets, and would certainly have no reason to do so. While we can easily imagine these creatures discovering the notion of temperature—the average kinetic energy of particles—it is hard to see how they could ever arrive at the notion of entropy. Hence the definition of entropy is not fully reductive; it *intrinsically involves the higher-level notion of a macrostate*. But macrostates, we saw, are defined by us. If their "size" depends on our description, so does their entropy. This dependence has led to the allegation that entropy (as construed in statistical mechanics) is in some way subjective or anthropomorphic.

Consider the following argument made by Jaynes:

> Entropy is an anthropocentric concept, not only in the well-known statistical sense that it measures the extent of human ignorance as to the microstate. *Even at the purely phenomenological level, entropy is an anthropocentric concept*. For it is a property, not of the physical system, but of the particular experiments you or I choose to perform on it. (Jaynes 1983, 86, emphasis in original)

The passage suggests an analogy between statistical mechanics and quantum mechanics, where dependence on the observer, or the experiment performed, is often asserted.[13] But we need endorse neither the analogy with quantum mechanics nor a subjective interpretation of probability in general to accept Jaynes's principal argument, which he states very clearly:

> Thermodynamics knows of no such notion as the "entropy of a physical system." Thermodynamics does have the concept of the entropy of a *thermodynamic* system; but a given physical system corresponds to many different thermodynamic systems. (Jaynes 1983, 85)

13. Jaynes (1983), 87. Conversations with Wigner, which Jaynes mentions, may have suggested this analogy, but in view of the fundamental differences between the two theories, it is problematic; see chapter 4.

We might object that the "real" entropy of a physical system is the entropy calculated by considering every single degree of freedom, an objection Jaynes overrules:

> There is no end to this search for the ultimate 'true' entropy until we have reached the point where we control the location of each atom independently. But just at that point the notion of entropy collapses, and we are no longer talking thermodynamics. (Jaynes 1983, 86)

The last sentence is meant to imply subjectivity, to assert that thermodynamics applies only as long as some ignorance remains. On my reading, however, the argument turns on individuation rather than ignorance. "Talking thermodynamics" calls for at least two levels of description, two schemes of individuation, so that the question of what fraction of the microstates corresponds to a certain macrostate makes sense. Our conscious molecule that recognizes only different microstates, but no higher-order states, would therefore indeed be unlikely to grasp the concept of entropy. Not because, as Jaynes suggests, its perfect knowledge of the microlevel would make it omniscient, whereas entropy requires a certain degree of ignorance, but because it would lack a suitable representation scheme.

This reading is confirmed by a comment Jaynes makes in reply to the criticism that his argument about the nature of entropy would apply just as well to energy:

> Not so! The difference is that energy is a property of the microstates, and so all observers, whatever macroscopic variables they may choose to define their thermodynamic states, must ascribe the same energy to a system in a given microstate. But they will ascribe different entropies to that microstate, because entropy is not a property of the microstate, but rather of the reference class in which it is embedded. As we learned from Boltzmann, Planck, and Einstein, the entropy of a thermodynamic state is a measure of the number of microstates compatible with the macroscopic quantities that you or I use to define the thermodynamic state. (Jaynes 1983, 78)

The difference Jaynes points to between entropy and energy is in line with the difference already noted between properties that are fully reducible to microproperties and entropy's pseudo-reducibility. Yet contra Jaynes, I distinguish between description dependence and full-blown subjectivity. Granted, we are the ones who define macrostates, but once a description has been specified, the entropy of the macrostate answering to that description is not subjective, but a matter of fact. The importance of the level at which a particular state is described, and its implications for the state's entropy, is illustrated in Sklar (1993, 346) by adducing an instance of Gibbs's paradox. Hydrogen molecules come in two forms: ortho-hydrogen, where the two protons have parallel spins, and para-hydrogen, where the spins are antiparallel. The question is whether, when the two kinds of hydrogen are mixed, the system's entropy increases. It stands to reason that if we consider the two quantities simply as hydrogen, there is no change in entropy, whereas if we distinguish between ortho- and para-hydrogen, and describe the process as mixing two different materials, the entropy increases.[14] The increase makes sense when we realize that *separating* mixed hydrogen into ortho- and para-hydrogen requires work, which is not the case when the difference between the two kinds is ignored, so that we are merely mixing two volumes of the same gas. We should therefore conclude, with Sklar, that although the difference in entropy between the two cases of mixing depends on the level of description, it is not a subjective difference. In other words, we can choose the level of detail we need and characterize our types accordingly, but the number of microstates realizing a certain type is not up to us.

Thus far, the case of statistical mechanics suggests that stability in this context is a distinctly macro phenomenon; even when all of a system's microstates are equally probable and equally unstable, it is clear that all its macrostates are not.[15] We will see in later chapters that this

14. For example, when we consider the two volumes as hydrogen, switching a molecule from each container to the other container does not make a difference, but when we distinguish between the ortho- and para-hydrogen, it constitutes a physical change.

15. Microstates too can, of course, vary in stability. The point is that even when they don't, macrostates may still vary.

difference between levels of description is characteristic of various other paradigmatic examples where privileged stable states emerge from an indistinguishable multitude of microstates. The case of statistical mechanics also shows that a system's temporal evolution cannot be accounted for by adducing probabilistic considerations without anchoring them in the system's underlying dynamics. Last, it suggests that a shift in perspective that introduces different concepts and categories is likely to reveal phenomena and laws unseen from the previous perspective. In this sense, the two levels of description are *conceptually* distinct; the most detailed description of the basic level fails to do the explanatory work needed if we are to understand the higher one. Furthermore, it is typically the case that understanding the higher level requires us to abstract from the details of the basic level and explain why they do *not* matter, that is, to show that some phenomena at the higher level are *insensitive* to them. In this sense, the conceptual independence of the higher level is accompanied by (a degree of) *causal* independence that can be expressed in the language of physics.[16]

3. The Origin of Directionality

We must now confront the third problem, the notorious problem of directionality. Recall that statistical mechanics was meant to explain irreversibility, that is, we sought to understand not only why higher-entropy states are more probable, but also—and these questions are not equivalent—why a system moves, or most probably moves, from less probable to more probable states rather than in the opposite direction. Here, the fact that the underlying dynamics, when taken to be correctly described by classical mechanics, are not only deterministic but also time-reversal symmetric, constitutes a formidable difficulty. The concern now is not merely, as it was in light of the aforementioned explanatory gap, that no basis for thermodynamic behavior can be found in the underlying dynamics, but the far more serious concern that the two are

16. To reiterate a point emphasized in previous chapters, on my broad picture of causation, causal explanation should include explanation of what does not happen and identification of factors that make no difference to what does happen.

in fact *incompatible*. Arguments for the incompatibility of irreversibility with classical mechanics were made early on by Loschmidt, Zermelo, and Poincaré, and have been widely discussed in the literature.[17] Loschmidt argued that the time-reversal symmetry of classical mechanics implies that for any possible entropy-increasing trajectory, there should be a possible entropy-decreasing one. Zermelo argued, on the basis of Poincaré's theorem, that given unlimited time, a classical system with a fixed number of particles and fixed energy, and confined to a specific space, will in general (i.e., with the exception of a measure zero set of initial conditions) repeatedly pass arbitrarily close to its starting point. In the long run, then, it yields recurrence rather than irreversibility. Boltzmann's response to these arguments was that the probabilistic construal of the second law does, in fact, render it compatible with the underlying dynamics.[18] For one thing, fluctuations are permitted, so entropy can indeed decrease as often as it increases, but the time spent by the system in any one macrostate will still be proportional to that state's probability, so fluctuations from equilibrium are unstable and short-lived. For another, if we assume low-entropy initial conditions, the probability that the system evolves into higher-entropy macrostates is overwhelming. Furthermore, without contesting Poincaré's theorem, the recurrence time it implies is so colossal that it is completely beyond human experience and thus generates no conflict with the shorter-range irreversible phenomena we do experience.

More recently, the reversibility objection has received a more general formulation: statistical mechanics employs two resources, the underlying mechanics and probability theory. The former is time-reversal symmetric, and the latter, being a mathematical theory, cannot create a temporal direction either. If these are the only resources available to statistical mechanics, there is no way to engender irreversibility. In other words, irreversibility seems incompatible not only with classical mechanics but also with the *combination* of mechanical and probabilistic arguments that were supposed to restore compatibility. To illustrate this problem, imagine a group of ice skaters at a large rink. In general, they

17. See Sklar (1993), Albert (2000), Uffink (2007).

18. I mention only those of Boltzmann's contentions that are relevant to my argument.

are more or less evenly scattered over the ice rink—the equilibrium state—but occasionally we observe fluctuations, for example, more skaters than usual concentrated in a small area of the rink. When encountering such a fluctuation, probabilistic considerations would lead us to assume that in a few seconds everything will go back to normal, but also that a few seconds *ago*, everything was perfectly normal and the skaters were evenly distributed. The analogous problem vis-à-vis statistical mechanics is that although it is true that probabilistic reasoning leads us to expect that, given a relatively low-entropy state, entropy is likely to increase going into the future, the same reasoning leads to the conclusion that entropy is equally likely to have been higher when we look toward the past. In other words, from such a relatively low-entropy state, entropy is just as likely to spontaneously decrease as it is to increase! The disconcerting conclusion of this analysis is that, unlike classical thermodynamics, statistical mechanics cannot account for irreversibility or time asymmetry.

There are different strategies for handling this problematic situation. One of them, already suggested by Boltzmann himself, is to assume that the conditions that actually obtained at the beginning of our universe were, as a matter of contingent fact, extreme low-entropy conditions. Trajectories that start off from such initial conditions are indeed highly likely to evolve into macrostates of higher entropy. Since this "past hypothesis" (Albert 2000) seems to have, as Russell remarked in a different context (1919, 71), "all the advantages of theft over honest toil," considerable effort has been put into justifying the special initial conditions as reasonable in light of our best cosmology, as "typical" in some technical sense, as the only ones compatible with the existence of human life, or, more simply, as those recorded in human memory and experience.[19] Another strategy introduces some sort of disturbance that *interferes* with the underlying time-symmetric dynamics (Albert 2011). The assumption here is that random "kicks" such as those mandated by the Ghirardi-Rimini-Weber (GRW) version of quantum mechanics, or

19. The latter solution, proposed in Hemmo and Shenker (2012c), is the most recent in this category. Further references to discussions of typicality apropos justification of the Lebesgue measure are mentioned in note 4 of this chapter; see also note 11.

such as might reach the system from the outside, jolt the system out of the trajectories dictated by the time-symmetric dynamics. The argument is that the effect of such random jolts differs in different cases: the overwhelming majority of trajectories are well behaved (i.e., move toward higher-entropy states), and would most probably be replaced, after being kicked, with trajectories that are likewise well behaved. On the other hand, trajectories moving toward lower-energy states are a small minority and, when perturbed, are likely to be replaced by trajectories belonging to the well-behaved majority, and to end up in higher-entropy states. The challenge for this approach is to find empirical support for the random perturbations in question and to demonstrate their differing effects on different trajectories. A third strategy introduces the desired time asymmetry as an additional assumption, sometimes deeming it a legitimate expression of the asymmetry of the causal relation. Ironically, this idea originates in a *critique* of Boltzmann's derivation of asymmetry. It was pointed out by Burbury ([1894] 1895) that by assuming the velocities of colliding particles to be independent prior to the collision but not afterward, Boltzmann tacitly presupposed the asymmetry he claimed to be proving.[20] This setback can, however, be turned into an advantage if the time-asymmetric independence assumption is justified in terms of the asymmetry of causation.[21] Critics of causation in general, and causal asymmetry in particular (e.g., Huw Price) will, of course, reject this solution.

All these approaches are based on causal assumptions: either the general assumption regarding causation's inherent asymmetry, or specific assumptions regarding specific physical processes that give rise to the stability and directionality familiar to us from experience.[22] The need to combine stability considerations with specific dynamical arguments will be emphasized in other places in this book. The emergence of sta-

20. How this epiphany played out is described in Price (1996), chap. 2.

21. See Penrose and Percival (1962) and Reichenbach (1956).

22. The strategies I have mentioned are not the only ones that have been proposed. An alternative strategy, less relevant to the questions I address in this chapter, and to the project of understanding causation, is to identify the direction of increasing entropy, by definition, with the direction of time.

bility characteristically involves a many-one relation that is sustained by the insensitivity of a certain type of state to differences between the many ways it can be realized. But adducing this structure, which is common to various systems that manifest stability, falls short of providing a full account of the actual processes by which stability is reached and maintained in any specific case. A full account, as we have seen, must anchor the formal structure in the relevant physical reality. This desideratum has important implications for the intertheoretic relationships of reduction and emergence, explored in chapter 7. Higher-level explanations may indeed involve conceptual novelties that reflect a certain degree of emergence and autonomy, attesting to the independence of higher levels from those on which they supervene. But unless this independence can also be given a detailed explanation in terms of the relevant physical theory, gaps and incompatibilities of the kind that plague statistical mechanics are likely to persist.

Statistical mechanics cautions us yet again to eschew the tendency to project a temporal direction onto the many-one relation, the recurrent formal structure I referred to in chapter 2 as emblematic of directionality. Stability has turned out to play an essential role in our understanding of the second law, a role that the deterministic laws of classical mechanics (or for that matter, the laws of quantum mechanics) could not play. Stability has thus secured its place as a distinct member of the causal family. Yet as we have seen, stability cannot, on its own, provide a complete account of directionality and irreversibility. Unless we make further assumptions (about causation or about a system's dynamics), the transition from unstable to stable states represented by the iconic many-one relation cannot serve as a full-blown arrow of temporal direction. Statistical mechanics also drives home the crucial significance of levels of description as reflecting the ways in which we carve up the world into physically relevant categories.

4

Determinism and Locality

CHAPTERS 2 AND 3 DISTINGUISHED DETERMINISM from neighboring concepts, including necessity, inevitability, and stability, with which it is often conflated. This chapter, by contrast, examines the relation between two causal constraints—determinism and locality—that at first glance appear independent, but prove to be linked by complex interconnections.[1] Clarifying these interconnections is particularly pertinent in the context of quantum mechanics (QM), which involves both indeterminism and nonlocality. In the philosophy of physics literature, the term *causality* usually refers to either determinism or locality. It refers to determinism in contexts involving the probabilistic laws of statistical mechanics and QM, or possible indeterminism in gauge theories, and refers to locality in contexts involving the special theory of relativity (STR), where locality is a fundamental constraint. Causality in the sense of locality is thus sometimes called *relativistic causality*.[2] In light of what has been recounted in the preceding chapters, such ambiguity regarding the notion of causation should not surprise us, but

1. An earlier version of this chapter (Ben-Menahem 2012) appeared in Ben-Menahem and Hemmo (2012). Substantial parts of that paper are included here with the kind permission of the publisher, Springer.

2. The philosophical literature in general (beyond the specific context of the philosophy of physics), focuses almost exclusively on causation in the sense of determinism. For instance, the Russell-Norton argument against causation, as we saw in chapter 2, actually targets determinism. Causality in the sense of locality is usually found only in philosophy of physics texts, as, e.g., in the title of Myrvold and Christian (2009): *Quantum Reality, Relativistic Causality, and Closing the Epistemic Circle*.

insofar as determinism and locality constitute distinct causal constraints, the conceptual relationship between them (and between indeterminism and nonlocality) merits consideration. Despite the importance of this question for a better understanding of causation, the philosophical literature has all but ignored it.

In what follows, I first consider the relation between determinism and locality in abstract terms, and then in the context of QM, where traditional conceptions of determinism and locality have been challenged. Limiting myself to the standard interpretation of QM, I compare the interpretative approaches of Schrödinger, Pitowsky, and Popescu and Rohrlich.[3] The common denominator of these approaches, which guided my choice, is that they focus on formal and conceptual, rather than dynamic, characteristics of QM. Although these authors do not address the relation between determinism and locality explicitly, their work provides important clues as to how it should be understood. I will argue that determinism and locality are independent concepts that nonetheless, under certain specific conditions (to be discussed below), offset each other, so that violation of the one permits satisfaction of the other. Indeed, this counterbalancing—or to put it differently, these payoff relations—will be the main thrust of this chapter.

Let me stress at the outset that although I begin by considering strict determinism and its relation to strict locality, the analysis extends to probabilistic correlations of the kind characteristic of QM. In fact, neither determinism nor locality need be understood as a binary notion; it is useful to replace strict determinism with quantitative assessments of degrees of correlation and to conceive of locality, too, as a matter of degree, so that theories can be more or less deterministic as well as more or less local.

Determinism having been analyzed in chapter 2, we can turn directly to locality. Locality has both a spatial and a temporal aspect; it asserts

3. There is no one "standard" interpretation, but the term is used here, as is common, to refer to descendants of the Copenhagen interpretation. Pitowsky's interpretation, e.g., is based on the Birkhoff–von Neumann axiomatization, the cornerstone of the standard interpretation. Rival interpretations, such as Bohm's, GRW, modal interpretations, and the many-worlds interpretation, which merit separate analysis, are not discussed here.

that physical interaction is spatially continuous and constrains the speed of physical interactions to exclude instantaneous propagation of impact and information.[4] Locality, like determinism, has classical origins, for example, the classical notion of contiguity and the idea that there are no "jumps" in nature. In contemporary physics, locality is generally taken to mean Lorentz invariance, and the upper bound on the speed of propagation is the speed of light. (The temporal asymmetry requirement—namely, the requirement that a cause precede its effect—is often added.[5]) Comparing the notion of locality with that of determinism, the two notions appear to be completely independent.[6] Locality entails that *if* there is a cause, it must act locally, that is, continuously and at a finite speed, but does not entail either that every event has a cause, or that the same cause must have the same effect. In the same vein, continuous and finite-speed interactions can be deterministic or indeterministic. The latter possibility describes the case where, despite an interaction's continuity and finite speed, there are no laws ensuring that recurrence of the initial conditions entails replication of the trajectory in its entirety. Conversely, deterministic interactions can, in principle, be continuous or discontinuous, instantaneous or of finite speed. Undoubtedly, actual theories familiar to us from the history of science may induce us to conjoin determinism and locality. The deterministic laws of Newtonian mechanics, for instance, have specific mathematical characteristics, such as analyticity, and hence also presuppose that physical interactions are spatially continuous, even if not of finite speed. We are thus accustomed to a conception on which physical processes and the laws describing them are both continuous and deterministic. This picture leaves no

4. Here I am referring to locality, not Bell-locality; see note 13 in this chapter.

5. See Frisch (2009b; 2014). As noted in chapter 1, this causal constraint, despite its importance, is not discussed in this book.

6. In antiquity and in the Middle Ages, however, they were not conceived as independent; see Glasner (2009), chap. 3. One reason for the divergence between ancients and moderns on this point is that in antiquity, determinism (though not under that name; see the beginning of chapter 2 above) was generally understood in terms of the universality requirement (every event has a cause) rather than the regularity requirement (same cause, same effect). On this construal, it is easier to appreciate why the contiguity of interaction, which excludes "jumps," would be taken to exclude chance, i.e., spontaneous occurrences.

room for spatial "jumps," but the legitimacy of infinite speed still allows for temporal "jumps." In other words, in classical physics, we ordinarily think in terms of a combination of determinism and spatial continuity, though not full-blown locality. Yet this combination is not forced on us by the concepts of determinism or locality per se. From the logical point of view, all four combinations seem conceivable:

1. locality and determinism
2. nonlocality and determinism
3. locality and indeterminism
4. nonlocality and indeterminism

Note, however, that the fourth combination, nonlocality and indeterminism, though logically possible, poses a serious epistemic difficulty. Nonlocality—that is, an instantaneous interaction between distant events, or a transmission of signals between them at a speed exceeding the speed limit—can only be demonstrated via the existence of recurring *correlations* between such distant events. Individual nonlocal interactions would not be identified by us as *interactions*, but seen, instead, as the occurrence of independent and causally unrelated events. But recurring correlations, even merely probabilistic ones, introduce at least some degree of regularity, or determinism, into the picture. A *local* indeterministic influence could perhaps still be identified as such when the actual trajectory of the action is visible. For example, if we observed that identical pushes of a ball result in its moving haphazardly in various directions, we could, perhaps, still think of the push as a cause, albeit a cause that does not act deterministically. By contrast, a *nonlocal* indeterministic interaction would in all likelihood escape our attention. Hence the fourth possibility, combining nonlocality and indeterminism, can appear in our theories only in a qualified form; such theories will not be totally indeterministic with regard to all parameters. Surprisingly, then, a grain of determinism turns out to be necessary, de facto, for nonlocality; it is necessary for the formulation of a nonlocal theory.[7]

7. Again, by "a grain of determinism," I do not mean that strict universal laws are necessary; probabilistic dependence would suffice for the detection of nonlocality. This is indeed what happens in QM.

Our actual theories thus should not, after all, feature absolute independence between determinism and locality.

The history of physics provides concrete examples of the possible combinations of locality and determinism. STR is local and deterministic, exemplifying (1) above, whereas Newtonian mechanics, which is nonlocal and deterministic, exemplifies (2).[8] Determinism is evidently not a sufficient condition for locality. But is it a necessary condition? To put it contra-positively, is indeterminism sufficient for nonlocality? Here again, from the purely conceptual point of view, the answer seems to be negative. In actual theories, however, QM in particular, the picture is more complicated. Let me review the situation briefly. In the context of QM, the threat to locality is raised by the phenomenon of quantum entanglement, manifested in long-distance correlations that are maintained even when the particles in the correlated states are separated by space-like intervals. Entanglement was identified by Schrödinger (Schrödinger 1935) and has been amply demonstrated by experiment.[9] As is well known, correlations pose a greater challenge to our causal intuitions than singular events do. Whereas random singular events seem to be conceivable, systematic correlations with no causal explanation seem inconceivable. Explanations of such correlations would invoke either the direct causal influence of one event on another, or a "common cause" acting on the system in question at an earlier time and generating the correlated states.[10] Common causes are said to "screen off" the dependence between correlated states, meaning that given the common cause, the states no longer appear interdependent.[11] If the

8. Counterexamples to determinism in Newtonian mechanics are presented in Earman (1986) and Norton (2007); see chapter 2 herein, note 8. In general, however, Newtonian mechanics countenances determinism without temporal locality, attesting to determinism's insufficiency for locality.

9. Issachar Unna has suggested to me (in correspondence) that in an unpublished 1927 paper (1927a), Einstein anticipated entanglement. The Einstein-Podolsky-Rosen (EPR) argument and Schrödinger's response to it render this claim somewhat doubtful. For commentary on the paper, see the introduction to vol. 15 of *The Collected Papers of Albert Einstein* (Princeton, NJ: Princeton University Press).

10. Reichenbach (1956), chap. 19.

11. To test for the existence of a common cause, we compare the conditional probability of

explanatory options of direct influence and common cause are exhaustive, entangled quantum states should likewise be seen as indicating either that entangled states exert direct, albeit nonlocal, influence on one another, or that a common cause is responsible for the linkage between them. Accepting the former option—nonlocality—appears to generate conflict between QM and STR, a theory committed to locality. The alternative—namely, explaining entanglement by adducing common causes (local hidden variables, as they are usually referred to) that predetermine the states and are responsible for their correlations—is thus more appealing.[12]

Various arguments, however, beginning with the discovery of Bell's inequalities and their violation by QM, are generally thought to preclude this alternative.[13] In response to the conundrum, the following distinction between types of locality (nonlocality) has been introduced. The nonlocal correlations exhibited by entangled states are tolerable and considered consistent with STR as long as they do not allow superluminal "signaling"—transmission of information—between the remote states. Causality in the sense of locality (as defined above) is

---

the joint event (given the putative common cause) with the product of the conditional probabilities of the individual events (given the putative common cause). When these probabilities are equal, i.e., when the conditional events are independent, the condition event is a common cause. In this case (equal conditional probabilities), we speak of factorizability (nonfactorizability in the case of nonequal probabilities). For an analysis of the relationship between factorizability and the existence of a common cause, see Chang and Cartwright (1993). It argues that since, in the probabilistic (indeterministic) case, factorizability is not, in general, a necessary condition for the existence of a common cause, the nonfactorizability of quantum distributions does not exclude the possibility that the EPR correlations have common causes. Chang and Cartwright go on to propose a common-cause model of EPR, but it requires discontinuous causal influences and is manifestly nonlocal.

12. In the context of discussions of Bell's inequalities, the assumption that a common cause (whether deterministic or stochastic) exists is sometimes referred to as locality, or Bell locality. Bell locality is not identical with locality as characterized above, as it is committed not only to the continuity and finite speed of any causal interaction *that might exist*, but also to the actual existence of a cause—a "screening-off" event.

13. There are, however, interpretations of QM, e.g., Bohm's theory, and the many-worlds interpretation, that reject this conclusion. As mentioned, this chapter assumes the standard interpretation. See also notes 17 and 29 in this chapter.

thus reduced to a constraint on signaling rather than on correlation in general.[14] On this understanding of locality, entangled states escape the dilemma: they result neither from nonlocal causal interaction, nor from preexisting common causes. The distinction between locality and no signaling constitutes a significant deviation from the traditional understanding of locality, on which the only possibilities were locality plus no signaling and nonlocality plus signaling. The distinction makes room for a new possibility previously taken to be incoherent: nonlocality plus no signaling. (The fourth possibility, locality plus signaling, remains incoherent.) Using this new terminology, QM can be said to satisfy locality, since despite entanglement, nonlocal signaling is prohibited. The advantage of this solution is obvious—there is no conflict with STR. But it has a significant drawback; not only are the strange correlations left unexplained, they are also deemed inexplicable, and indeed, not even in need of explanation.[15]

The distinction between (legitimate) nonlocal correlations and (illegitimate) signaling would not amount to a genuine difference were it the case that the legitimate nonlocal correlations could be used for signaling.[16] But as it happens, they cannot! Here indeterminism plays a

14. The notion of signaling is tinged with anthropomorphism, but I will not attempt to refine it. No signaling is not identical with Lorentz invariance; a theory can prohibit signaling without being Lorentz invariant.

15. See Redhead (1987) and Maudlin (1994) for finer distinctions between the various notions of locality. Although the distinction between correlation and signaling is widely accepted, many physicists and philosophers contend that nonsignaling correlations nevertheless violate, as Rohrlich puts it, the spirit, if not the letter, of STR.

16. The threat of action at a distance had already been noted by Einstein in 1927. He illustrated the problem by a thought experiment involving a photon that, after hitting a semitransparent mirror, is in superposition of a reflected wave and a transmitted wave. A measurement that detects the photon in one of these states immediately destroys the superposition, affecting the other part of the wave, regardless of how distant it is. Commenting on this thought experiment in his 1929 Chicago lectures, and responding to the concern about QM's inconsistency with STR, Heisenberg asserts: "The experiment at the position of the reflected packet thus exerts a kind of action (reduction of the wave packet) at the distant point occupied by the transmitted packet, and one sees that this packet is propagated with a velocity greater than that of light. However, it is also obvious that this kind of action can never be utilized for the transmission of signals so that it is not in conflict with the postulates of the theory of relativity" (Heisenberg 1930, 39).

crucial role. To appreciate the intricate dance between indeterminism and locality, note that the very idea of separating correlation and signaling seems paradoxical; correlation is certainly necessary for signaling, and at first glance, also seems sufficient. How, then, can we conceive of correlations between systems that satisfy the no-signaling constraint? It turns out that entangled states are prevented from signaling—that is, cannot serve to transmit information—because they are not predetermined! Had the results of measurement been predetermined, the experimenter at one end of the entangled system could, by looking at her results, immediately know whether the experimenter at the other end had made a measurement that interfered with the predicted outcome.[17] In the absence of such predetermination, even though the results at one end are correlated with those at the other, they do not disclose information about them. In other words, it is the *indeterminism* of QM (on the standard interpretation) that eliminates the possibility of signaling and thereby saves QM's consistency with STR. What an ironic twist of Einstein's vision!

I argued above that nonlocality cannot be identified as such if the nonlocal correlations are completely random. Nonlocal theories must therefore accommodate correlations, that is, must countenance at least some measure of determinism. In the case of QM, it is indeed the correlations exhibited by entangled states (to wit, not only strictly deterministic correlations) that suggest nonlocality. Yet we now realize that

---

This may be the first attempt to distinguish nonlocality from signaling so that a nonlocal theory could still be consistent with STR. However, Heisenberg only addressed the immediate collapse of the wave function at a distance (i.e., at a distance from the location of the measurement), not entanglement in general. The distinction did not convince Einstein, who continued to worry about nonlocality and the consistency problem it gives rise to. The distinction between locality and no signaling gained prominence after Bell (1954, 1966), especially when QM's nonlocal correlations were confirmed by experiment. The compatibility of nonlocality and no signaling is supported by the no-signaling theorem; see, e.g., Cushing (1994), appendix 2 to chap. 10.

17. Bohmian QM seems to provide a counterexample, since although deterministic, it does not allow signaling. Recall, however, that in Bohmian QM the equilibrium state excludes knowledge of the predetermined states. In the absence of this information, the experimenter cannot use the correlations for signaling.

satisfaction of the no-signaling constraint is made possible by the fact that QM also accommodates (at least a measure of) *in*determinism. The upshot of these twin considerations is that theories such as QM (and there could be a family of such theories), which allow for entanglement but exclude signaling, strike a delicate balance between determinism and indeterminism.

Linking determinism and locality to a central tenet of QM—the Heisenberg uncertainty relations—sheds light on how this balance is maintained in QM. Whereas the connection between the uncertainty relations and indeterminism is salient, their connection to nonlocality and entanglement is far less obvious. There are nonetheless strong arguments to the effect that not only indeterminism, but also entanglement and nonlocality, are implicit in and mediated by the uncertainty relations. Similarly, there are arguments showing that the combination of nonlocality and no signaling entails uncertainty. I will discuss three approaches that link uncertainty and nonlocality: the first two, due to Schrödinger and Pitowsky, argue from the uncertainty relations to entanglement, the second, following Popescu and Rohrlich, from entanglement to the uncertainty relations.

1. Schrödinger's Approach

Given Schrödinger's critical stance vis-à-vis the Copenhagen orthodoxy, his "cat" paper—"The Present Situation in Quantum Mechanics" (1935)—is an admirable attempt to provide an unbiased account of the theory.[18] Although some of its arguments (first and foremost the cat paradox) have been used by critics of the standard interpretation to find fault with this interpretation and its nonintuitive implications, the paper actually drives home QM's distinctly nonclassical nature, thereby supplying a firm basis for the standard interpretation. (Unfortunately, Schrödinger's reputation as an adversary of the standard interpretation tends to mask the message of this important paper.) The nonclassical

18. Schrödinger's interpretation, particularly its relation to the Pusey-Barrett-Rudolph (PBR) theorem (Pusey, Barrett, and Rudolph 2012), is also discussed in Ben-Menahem (2017).

nature of QM is epitomized by the uncertainty relations, which, Schrödinger maintains, are the key to interpreting quantum states and quantum probabilities. In restricting the determinacy of certain pairs of basic physical parameters, "the classical notion of *state* becomes lost in that at most a well-chosen half of a complete set of variables can be assigned definite numerical values" (1935, 153). Remarkably, Schrödinger does not take this to be a merely epistemic problem. He does not see the uncertainty relations as limiting only what we can know or measure, but as pertaining to the *existence* of determinate states. The very *assignment of definite values* to all variables, he asserts, is excluded by the uncertainty relations. Schrödinger therefore rules out the possibility that quantum probabilities and uncertainties are analogous to probabilities in statistical mechanics, reflecting human ignorance rather than genuine indeterminacy of quantum states.

Moreover, evaluating the implications of the uncertainty relations, Schrödinger identifies the structure of the event space of QM as the basic feature that distinguishes quantum from classical mechanics.[19] Whereas many of his colleagues emphasized the difference between the deterministic character of classical mechanics and the probabilistic nature of QM, Schrödinger emphasizes the *nonclassical nature of quantum probability*, an insight that was confirmed decades later by the work of Bell, Gleason, Kochen and Specker, and others. (The nonclassical nature of quantum probability is the core of Pitowsky's approach, discussed in the next section.)

> One should note that there was no question of any time-dependent changes. It would be of no help to permit the model to vary quite "unclassically," perhaps to "jump." Already for the single instant things go wrong. At no moment does there exist an ensemble of classical states of the model that squares with the totality of quantum mechanical statements of this moment. The same can also be said as follows: if I wish to ascribe to the model at each moment a definite

19. Although Schrödinger was, in general, committed to continuity, and very critical of the idea of "quantum jumps" (1952), he does not take the positing of discrete states to be the main difference between quantum and classical physics.

> (merely not exactly known to me) state, or . . . to *all* determining parts definite (merely not exactly known to me) numerical values, then there is no supposition as to these numerical values *to be imagined* that would not conflict with some portion of quantum theoretical assertions. (Schrödinger 1935, 156, emphasis in original)

The import vis-à-vis indeterminism is straightforward: "If even at any given moment not all the variables are determined by some of them, then of course neither are they all determined for a later moment by data obtainable earlier" (154). And further, "if a classical state does not exist at any moment, it can hardly change *causally*. What do change are the . . . probabilities, *these*, moreover, causally" (154, emphasis in original).

Schrödinger takes the Ψ function to represent a maximal catalog of possible measurements. It embodies "the momentarily attained sum of theoretically based future expectations, somewhat as laid down in a *catalog*. It is the . . . determinacy bridge between measurements and measurements" (1935, 158, emphasis in original). As such, with each new measurement, the Ψ function undergoes a change that "*depends on the measurement result obtained*, and so *cannot be foreseen*" (158, emphasis in original). The catalog's maximality, or completeness—a consequence of the uncertainty relations—entails that we cannot have two Ψ functions of the same system, one of which is included in the other. "Therefore, if a system changes, whether by itself or because of measurements, there must always be statements missing from the new function that were contained in the earlier one" (159). In other words, any additional information arrived at by measurement must change the previous catalog by *deleting* information from it. This is the gist of the "disturbance" that ensues from measurement. True statements that were part of the catalog prior to the measurement become false. This means that at least some of the previous values have been destroyed.[20]

20. This explanation of disturbance anticipates Spekkens's derivation of disturbance from the principle he dubs the "knowledge-balance principle," according to which, in a state of maximal knowledge, the amount of knowledge is equal to the amount of uncertainty, that is, the number of questions that can be answered about a system's physical state is equal to the number of questions that cannot be answered (Spekkens 2005).

Thus far Schrödinger has used the uncertainty relations to derive three features of QM: indeterminism, "disturbance" by measurement, and the deviation of quantum probabilities from classical probability theory. As I noted, only the first two were fully acknowledged at the time. Now comes entanglement. This new feature, he shows, also follows from the maximality or completeness of the Ψ function, that is, from the uncertainty relations. He argues as follows. A complete catalog of two separate systems is, ipso facto, also a complete catalog of the combined system, but the converse does not follow. "*Maximal knowledge of a total system does not necessarily include total knowledge of all its parts, not even when these are fully separated from each other and at the moment are not influencing each other at all*" (Schrödinger 1935, 160, emphasis in original). The reason we cannot infer such total information is that the maximal catalog of the combined system may contain conditional statements of the form: *if* a measurement on the first system yields the value x, a measurement on the second will yield the value y, and so on. He sums up: "Best possible knowledge of a whole does not necessarily include the same for its parts. . . . The whole is in a definite state, the parts taken individually are not" (161). In other words, separated systems can be correlated or entangled via the Ψ function of the combined system, but this does not mean that their individual states are already determined! Schrödinger's argument clarifies the conclusion reached above regarding the merits of combining determinism—to detect nonlocality—and indeterminism—to prevent signaling. A degree of determinism is supplied by conditional statements, derived from conservation laws, that generate the correlations. Indeterminism pertains to the individual outcomes.

Schrödinger's argument is purely conceptual. As we have just seen, he does not examine entanglement as a physical process in space and time (such processes would be described as intuitive or *anschaulich* in the idiom of the day), but rather as a conceptual possibility emerging from the uncertainty relations and the notion of a maximal catalog. Similarly, he does not construe the collapse of the wave function as a physical process. It too is a formal property of the Ψ function, a function that is in any event situated in configuration space, not real space.

The only dynamical consideration that figures in Schrödinger's argument is that to be entangled, two systems must have interacted in the past. Entanglement cannot be generated between separated systems.

Schrödinger tells us (1935, n. 7) that his paper was written in response to the Einstein-Podolsky-Rosen (EPR) paper published earlier that year (Einstein, Podolsky, and Rosen 1935). Schrödinger is usually thought of as Einstein's ally in opposing the Copenhagen interpretation, and, indeed, Schrödinger and Einstein often shared their misgivings about it. It is therefore easy to overlook the fact that in this paper, Schrödinger, without saying so explicitly, is critical of Einstein's position. The paper puts forward a more lucid and effective critique of the EPR argument than that voiced in Bohr (1935). The EPR argument purports to show that the correlations between the remote parts of a system—the conditional statements—entail that each individual state has a determinate value *prior* to measurement. Schrödinger points out, first, that such determinacy is precluded by the uncertainty relations, properly understood, and second, that, given his reading of the Ψ function as a maximal catalog of possible measurements, the indeterminacy of individual outcomes makes perfect sense. In other words, the EPR argument seeks to reveal the existence of predetermined states underlying the correlations, which amounts to understanding them in terms of common causes, but Schrödinger realizes that this solution does not work, and suspects that QM may be incompatible with STR. This concern could only be addressed after Bell, by invoking the nonlocality/no-signaling distinction. On the epistemic approach, to which we now turn, however, this concern is altogether moot.

2. Pitowsky's Approach

In his "Quantum Mechanics as a Theory of Probability" (2006), Pitowsky further elaborates the axiomatic approach originating with Birkhoff and von Neumann. Building on their classic axiomatization in terms of the Hilbert space structure of quantum events and its relation to projective geometry, Pitowsky seeks to incorporate later developments, such as Gleason's theorem (1957), and Bell (and Bell-type) inequalities, and

identify their roots in the axiom system. The 2006 article wraps up much of Pitowsky's earlier work on the foundations of QM, focusing, in particular, on the nonclassical nature of quantum probability. The ramifications of the nonclassical structure of the quantum probability space, he argues, include indeterminism, loss of information upon measurement, entanglement, and Bell-type inequalities. The ramifications are also closely linked to Gleason's theorem (1957), Kochen and Specker's theorem (1967), and, as shown in Bub and Pitowsky (2010), to the information-theoretic no-cloning (or no-broadcasting) principle. I will mention only those aspects of Pitowsky's work that enhance our understanding of the relation between locality and determinism.

The nonclassical nature of quantum probability manifests itself in the violation of basic classical constraints on the probabilities of interrelated events and is reflected in simple paradigm cases such as the two-slit experiment. For example, in classical probability theory, it is obvious that if we have two events E_1 and E_2 with probabilities p_1 and p_2, and their intersection, whose probability is $p_{12} = p_1 \cdot p_2$, the probability of the union $(E_1 \cup E_2)$ is $p_1 + p_2 - p_{12}$, and cannot exceed the sum of the probabilities $(p_1 + p_2)$.

$$0 \leq p_1 + p_2 - p_{12} \leq p_1 + p_2 \leq 1$$

In the two-slit experiment, however, the predictions of quantum mechanics violate this classical constraint, as there are areas on the screen that get more hits when the two slits are open together for a time interval Δt, than when each slit is open separately for the same interval Δt. In other words, contrary to the classical principle, we get a *higher* probability for the union than for the sum of the probabilities of the individual events. (Since the violation emerges from comparing different experiments—different samples—it does not constitute an outright logical contradiction.) This phenomenon is usually described in terms of interference, superposition, wave-particle duality, nonlocal influence of one open slit on particles passing through the other, and so on. Pitowsky's point is that before we venture to suggest theoretical *explanations* of the observations predicted by QM (and confirmed by

experiment), we must acknowledge the bizarre nature of the phenomenon itself—nothing less than violation of a highly intuitive principle of classical probability theory.

QM predicts analogous violations of other classical rules of probability. The most famous such violation is the violation of Bell's inequalities, which, like the above rule, can be derived from classical combinatorial considerations. Pitowsky (2006 and references therein) showed that from the analogue of the above classical rule for three events it is just a short step to Bell's inequalities, equalities that are known to be violated by QM (and experiment). The inequalities are thus directly linked by Pitowsky to Boole's classical "conditions of possible experience." What makes the violated conditions "classical" is the underlying assumption that the entities in question have determinate, measurement-independent properties; just as balls in an urn are red or wooden, to derive Bell's inequalities it is assumed that particles have a definite polarization, or a definite spin in a specific direction, and so on. The violation of the classical principles of probability compels us to renounce this classical assumption and replace it with a new understanding of quantum states and quantum properties. What does it mean for a particle to *be* in a certain state, say, spin-1 in the *x* direction, and what is the role of measurement in revealing this state? More generally, what is the meaning of the quantum state function? Pitowsky's answer is similar to that given by Schrödinger: the quantum state function is fundamentally different from a classical state, which represents physical entities and their properties prior to measurement. The quantum state function only keeps track of the *probabilities* of measurement results, and hence is a "book-keeping" device, as Pitowsky puts it (2006, 214), or as Schrödinger put it, a "catalog of possible measurements."

This understanding of the state function led Pitowsky to two further observations. First, in contrast to Schrödinger, he interpreted the book-keeping picture subjectively: quantum probabilities are understood as degrees of partial belief. Second, concurring with Schrödinger, he took the notorious collapse problem to be less formidable than it was on a realistic construal of the state function, for if what collapses is not a real entity in physical space, then there is no reason why the collapse should

be construed as a real physical process satisfying locality and Lorentz invariance.[21] There is thus a direct link between Pitowsky's taking QM to be primarily a theory of nonclassical probability and his renouncing what he and Bub, in a joint paper (2010), dubbed "two dogmas" of the received view—namely, the reality of the state function and the need for a dynamic account of the measurement process.

Like Schrödinger, Birkhoff, and von Neumann before him, Pitowsky takes the uncertainty relations to be the crucial feature demarcating the quantum domain from the classical (Pitowsky 2006, 214). The only nonclassical axiom in the Birkhoff–von Neumann axiomatization, and thus the logical anchor of the uncertainty relations, is the axiom of *irreducibility*.[22] Whereas a classical probability space is a Boolean algebra where for all events x and z:

$$x = (x \cap z) \cup (x \cap z^{\perp}) \text{ (reducibility)}$$

in QM, we get irreducibility, i.e. (with 0 as the null event and 1 the certain event):

$$\text{If for some } z \text{ and for all } x,\ x = (x \cap z) \cup (x \cap z^{\perp}), \\ \text{then } z = 0 \text{ or } z = 1$$

Irreducibility signifies the non-Boolean nature of the algebra of possible events, since the only irreducible Boolean algebra is the trivial

21. Pitowsky's epistemic interpretation of quantum states is often conflated with instrumentalism. The crucial difference between these positions is explicated in Ben-Menahem (2017). An insightful analysis of Pitowsky's interpretation can be found in *Theories*, chap. 4, by Bill Demopoulos (Harvard University Press, forthcoming). There are a number of other epistemic interpretations of quantum probabilities that address the measurement problem (and the issue of nonlocality, discussed in the following sections) along similar lines. See, in particular, the literature on QBism, for example, Fuchs (2014).

22. I am ignoring the minor differences between Pitowsky's formulation and that of Birkhoff and von Neumann. In comparison with previous axiomatizations, Pitowsky's treatment of the representation theorem for the axiom system, and in particular, his discussion of Solér's theorem, is a significant advance. The theorem, and the representation problem in general, are crucial for the application of Gleason's theorem, but need not concern us here.

one $\{0, 1\}$. As Birkhoff and von Neumann explain, irreducibility means that there are no "neutral" elements z, $z \neq 0$ $z \neq 1$ such that for all x, $x = (x \cap z) \cup (x \cap z^{\perp})$. (Were there such "neutral" events, we would have nontrivial projection operators commuting with all other projection operators). Intuitively, irreducibility embodies the uncertainty relations: when x cannot be represented as the union of its intersection with z and its intersection with $z^{\perp}$ (the complement of z), then x and z cannot be assigned definite values at the same time. Thus whenever $x \neq (x \cap z) \cup (x \cap z^{\perp,})$, x and z are incompatible, and consequently, measurement of one of them yields no information about the other. The axiom further implies genuine uncertainty—probabilities strictly between (unequal to) 0 and 1. In other words, it implies indeterminism. This result follows from a theorem Pitowsky calls the *logical indeterminacy principle*. It proves that for incompatible events x and y:

$$p(x) + p(y) < 2$$

The loss of information upon measurement—the phenomenon called "disturbance" by the founders of QM—also emerges as a formal consequence of the probabilistic picture.

Having shown that the axiom system entails genuine uncertainty, Pitowsky moves on to demonstrate the violation of the Bell inequalities—namely, the phenomena of entanglement and nonlocality. These violations already appear in finite-dimensional cases, and follow from the probabilities of the intersection of the subspaces of the Hilbert space representing the (compatible) measurement results at the two ends of the entangled system. Pitowsky shows, in both logical and geometric terms, that the quantum range of possibilities is indeed larger than the classical range, so that we get *more* correlation than is allowed by the classical rules; that is, we get nonlocality.[23] Whereas the usual response to this phenomenon consists in attempts to discover the dynamic that makes it possible, Pitowsky emphasizes his argument's

23. In *Quantum Probability, Quantum Logic Lecture Notes in Physics 321,* Pitowsky (1989) explores the geometric meaning of Boole's classical probability rules.

logical-conceptual nature, which renders it independent of specific physical considerations beyond those that follow from the non-Boolean nature of the event structure. He asserts:

> Altogether, in our approach there is no problem with locality and the analysis remains intact no matter what the kinematic or the dynamic situation is; the violation of the inequality is a purely probabilistic effect. The derivation of Clauser-Horne inequalities . . . is blocked since it is based on the Boolean view of probabilities as weighted averages of truth values. This, in turn, involves the metaphysical assumption that there is, simultaneously, a matter of fact concerning the truth values of incompatible propositions. . . . From our perspective the commotion about locality can only come from one who sincerely believes that Boole's conditions are really conditions of possible experience. . . . But if one accepts that one is simply dealing with a different notion of probability, then all space-time considerations become irrelevant. (Pitowsky 2006, 231–32)

Recall that in order to countenance nonlocality without violating STR, the no-signaling constraint must be satisfied. Since Pitowsky construes nonlocality in formal terms—as a manifestation of the quantum mechanical probability calculus, uncommitted to any particular dynamic—it stands to reason that the no-signaling principle can likewise be derived from probabilistic considerations. Indeed, it turns out that no signaling can be construed as an instance of a more general principle—the noncontextuality of measurement (Barnum et al. 2000). In the spirit of the probabilistic approach to QM, Bub and Pitowsky therefore maintain that no signaling "is not specifically a relativistic constraint on superluminal signaling. It is simply a condition imposed on the marginal probabilities of events for separated systems, requiring that the marginal probability of a B-event is independent of the particular set of mutually exclusive and collectively exhaustive events selected at A, and conversely" (2010, 443).[24]

24. In the literature, following Jarrett (1984) in particular, it is customary to distinguish outcome independence, which QM violates, from parameter independence, which it satisfies,

Pitowsky's formal approach takes both indeterminism and nonlocality to be embedded in the event structure of QM. It is a basic tenet of this approach that QM has no deeper foundation than this formal structure. Once we accept QM as a new and nonclassical theory of probability (or information), the argument goes, the intriguing problems of how nonlocal correlations arise, why measurement generates disturbance, and so on, can be set aside. Bub and Pitowsky draw an analogy between their formal approach to QM and Minkowski's geometric construal of STR, according to which relativistic effects such as the contraction of rods (in the direction of motion) and time dilation are kinematic effects of the geometry of spacetime that need no further explanation. Whether this analogy does indeed dispel the unsettling aspects of nonlocality is debatable, but whichever side we take in this debate, with regard to the relation between determinism and locality,[25] Pitowsky's approach yields the same conclusion as Schrödinger's. It is the uncertainty relations and the indeterminism they engender that enables nonlocal theories such as quantum mechanics to satisfy the no-signaling constraint.

---

a combination that enables the peaceful coexistence with STR. The noncontextuality of measurement amounts to parameter independence. But see Redhead (1987) and Maudlin (1994), among others, for a detailed exposition, and critical discussion, of the outcome independence–parameter independence distinction and its implications for QM's compatibility with STR.

25. In this debate, each side sticks to its intuitions. I believe that we should aspire to show how the logical-mathematical-probabilistic laws we take to be true are backed by physical principles. Bub and Pitowsky deem such a physical anchor unnecessary. Strong support for the former position—the view that physical grounding is needed—is provided by Bell's "How to Teach Special Relativity" (1976). As long as the controversy is limited to the question of what constitutes explanatory force in physics, it can, perhaps, remain unresolved, but with respect to the tension between QM and STR, the problem is more pressing. The fact that entanglement and nonlocality can be described in terms of information-theoretic constraints is indeed eye-opening, but it is not the whole story. To complete the story, it should also be told in the language of STR, where the structure of Minkowski spacetime is indispensable, and in the language of the physical processes that take place in this spacetime. There is thus no choice, in my view, but to bring entanglement back into space and time, and address the abovementioned worries about the coherence of our overall picture of the physical world.

3. Popescu-Rohrlich Approach

We have seen that according to both Schrödinger and Pitowsky, a formalism that incorporates the uncertainly relations (or, equivalently, incorporates the axiom of irreducibility and posits incompatible events), gives rise to nonlocality. The question that must now be considered is whether we can also move in the opposite direction, that is, whether quantum entanglement and nonlocality yields the uncertainty relations. A series of papers by Popescu and Rohrlich (Popescu and Rohrlich [1994, 1998]) sheds light on this intriguing question.[26] The original question addressed in these papers was whether the nonlocal correlations of QM could be tampered with or destroyed by a third party. The idea here is that if nonlocal correlations reflect superluminal communication between distant systems, it might be possible to interfere with this mysterious communication channel. To test this idea, Popescu and Rohrlich contemplated Jim the Jammer, who is situated in a position that enables him to jam the EPR correlations between Bob and Alice. As noted, nonlocality in itself can be countenanced as long as it does not entail superluminal signaling. Accordingly, the envisaged jamming must be such that Bob and Alice will not notice it.

Popescu and Rohrlich sought a quantitative assessment of the relation between nonlocality and no signaling; they tried to ascertain the *maximal* amount of nonlocality that does not lead to signaling, that is, the maximal nonlocality compatible with STR. Their initial conjecture was that the constraints of maximum nonlocality and no signaling suffice to recover QM precisely, no more, no less. Intuitive support for this conjecture came from the observation, already noted, that an indeterministic theory could be both nonlocal and consistent with STR. Hence the feasibility of the idea that QM strikes exactly the right balance between nonlocality, no signaling, and indeterminism. Are nonlocality and no signaling, then, sufficient to generate QM? Surprisingly, Rohrlich

26. The Popescu-Rohrlich approach was partly inspired by Aharonov's ideas about the relationship between nonlocality and indeterminism. These ideas were first presented in talks and classes, but are explicit in Aharonov and Rohrlich (2005, 85–87).

and Popescu answered in the negative: nonlocality plus no signaling spans a *family* of theories that includes, in addition to QM, a range of theories that are *more nonlocal* than QM.[27] All members of this family feature uncertainty relations analogous to those of QM, though possibly differing from them in the value of the numerical limit they set.[28] Here both (non)locality and (in)determinism have become quantitative, rather than binary, notions. Moreover, they have been shown to be mutually interdependent. The combination of nonlocality and no signaling is linked to, and made possible by, the uncertainty relations and the indeterminism they give rise to.

The Popescu-Rohrlich argument suggests that indeterminism is at least (part of) a sufficient condition for the peaceful coexistence of nonlocality and STR. On the face of it, the stronger claim that indeterminism is also a necessary condition for this coexistence seems unwarranted in light of Bohm's theory, which, despite being deterministic, does not allow signaling.[29] However, given that in Bohm's theory, due to the equilibrium conjecture, the predetermined states are unknown to the experimenter, and thus useless for signaling, this counterexample may turn out to be deceptive.[30] If so, a kind of epistemic indeterminism (such as is found even in Bohm's theory) is not only a sufficient condition for the peaceful coexistence of nonlocality and no signaling but also a necessary one.

The interconnection between nonlocality and indeterminism is further supported by Oppenheim and Wehner (2010), who argue that the two basic features of QM, nonlocality and the uncertainty relations,

27. In the Clauser-Horne-Shimony-Holt inequality, the classical limit reached by local realist considerations is $-2 \leq S \leq 2$. In QM this inequality can be violated, but as Boris Tsirelson (1980) has shown, there is an upper bound to this violation: $-2\sqrt{2} \leq S \leq 2\sqrt{2}$. Popescu and Rohrlich (1994), and Lo, Spiller, and Popescu (1998) show that the Tsirelson bound can be violated without violation of STR, that is, without violation of the no-signaling requirement.

28. In QM, for incompatible variables constrained by the uncertainty principle, such as a particle's position x and momentum p, the principle sets the limit $\Delta x \Delta p \geq h/4\pi$. In other nonlocal theories, the value of the limit may differ from the quantum mechanical limit.

29. Although it does not allow signaling, Bohmian QM is not Lorentz invariant; see Albert (1992, chap. 7).

30. See note 17 in this chapter.

"are inextricably and quantitatively linked" (1072), so that QM could not be more nonlocal than it is without violating the uncertainty principle. From a different perspective, Goldstein and colleagues (2009), in critiquing Conway and Kochen's "free will theorem" (2006), also stress the difference between deterministic and stochastic theories insofar as satisfaction of the no-signaling constraint is concerned. Their argument is particularly significant in view of Tumulka's relativistic version of the GRW theory (2006). Were Tumulka's argument to apply to a deterministic analogue of the GRW theory, we would have a deterministic Lorentz invariant version of QM, and hence a counterexample to the argument I have made in this chapter. Goldstein and colleagues show, however, that indeterminism plays a crucial role in Tumulka's argument.

Delicate payoff relations between determinism and locality also surface in attempts to explain the Aharonov-Bohm effect (1959). Classical electromagnetic theory (as represented in Maxwell's equations) involves both electric and magnetic fields, and their potentials. From a classical point of view, the fields are understood as physically real, whereas the potentials, which are underdetermined by the corresponding fields, are seen as "gauge dependent," that is, as part of the mathematical apparatus, and thus as devoid of physical significance. The Aharonov-Bohm effect challenges this classical picture. The effect consists in the phase shift of particles moving through a field-free region, suggesting either that the physical information about the system is not exhausted by the fields, or that fields *outside* the field-free region act nonlocally on the particles traveling within that region. In one version of the Aharonov-Bohm effect, a uniform magnetic field is generated inside a solenoid by turning on a current that runs through the solenoid. While the magnetic *field* is confined to the space inside the cylindrical solenoid, and vanishes (or is negligible) outside, the magnetic *potential* outside the solenoid is nonzero. Aharonov and Bohm (1959) showed that according to QM, when the current is on, the wave function of particles traveling through the field-free region (i.e., outside the solenoid) will undergo a phase shift (detectable by interference), a prediction later confirmed by experiment. Interpretations of this effect vary.

Aharonov and Bohm took it to demonstrate the physical meaning of the electromagnetic potentials, concluding that these gauge-dependent quantities do indeed have physical reality. On this interpretation, the potentials act locally, but, as a consequence of underdetermination, indeterministically. Alternatively, the effect has been understood as preserving determinism while illustrating nonlocality—the nonlocal influence of the field inside the solenoid on particles traveling outside it. The question of whether these interpretations are empirically equivalent or can be distinguished by experiment is still being debated.[31] The lessons of the Aharonov-Bohm effect for the conceptual relations between determinism and locality are therefore not as definite as those of quantum entanglement, but the debate nonetheless suggests a more complex interrelation than the initial impression of independence led us to assume.

To conclude, we have seen that in QM, determinism and locality (indeterminism and nonlocality) stand in complex payoff relations. QM demonstrates that it is indeterminism that makes possible the combination of nonlocality and no signaling. In theories that, like QM, permit nonlocal correlations, nonlocality and indeterminism "cooperate" to prevent signaling and protect compatibility with STR. The uncertainty relations thus play a major role in maintaining this cooperation.

31. For these and other interpretations, see Aharonov and Rohrlich (2005), Belot (1998), Healey (2007), Wu and Yang (1975), and Vaidman (2012). Aharonov no longer accepts the conclusion of the original Aharonov-Bohm paper (1959)—the reality of the potential. Instead, he endorses the view that the field has nonlocal influence (Aharonov and Rohrlich 2005, 87). This nonlocal influence, like entanglement, does not permit signaling, and in this respect provides further support for the payoff argument presented in this chapter.

5

Symmetries and Conservation Laws

FOR THOSE WHO PONDER the "unreasonable effectiveness of mathematics in the natural sciences," as Eugene Wigner (1960) did,[1] the many applications of the notion of symmetry, and the tremendous work this notion does for the physicist, certainly provide some of the most striking examples.[2] In the case of symmetry considerations, it seems, we don't merely use mathematical language to express familiar, or conjectured, physical laws, but we actually import some segment of mathematics into physics, and then use it to derive new physical laws. More than other empirical laws, symmetries appear to have an element of aprioricity that endows them with a special grace and nobility; they belong to the nomic aristocracy, as it were. Hermann Weyl went even further: "As far as I can see, all a priori statements have their origin in symmetry" (1952, 126). And though we now know that at least some—and conceivably all—of the symmetries of physics are not, in fact, a priori, they have not completely lost their privileged status despite the recent tendency toward nomic egalitarianism.

Symmetries do indeed underscore questions regarding the relation between a mathematical structure and its physical realization(s). As we will see in more detail in this chapter, we are occasionally confronted with the existence of what Michael Redhead (2003, 128) dubbed "sur-

1. See also Steiner (1998), which is a detailed exploration of the relation between mathematical structures and physical theories.

2. An earlier version of this chapter appeared in *Iyyun* 61 (2012): 193–218. It is included here with the journal's permission.

plus structure," where the same physical structure can be correlated with several distinct mathematical structures, giving us more freedom than we would like to have. This freedom creates a gap between the mathematical and physical realms, upsetting the correspondence that is generally expected to obtain between them. What physicists do in such cases, beginning with Einstein's famous "hole argument" (*Lochbetrachtung*)[3] and continuing with gauge theories, is impose *new symmetries* on the mathematical side, that is, impose equivalence relations that construe several mathematical states as the same physical state. This procedure eliminates the freedom generated by the surplus structure and restores the tight fit between the mathematical and physical realms.

This chapter focuses on the place of symmetry in the network of causal constraints. It argues that symmetry principles play a causal role on a par with that of other causal constraints, and examines some of the interconnections between symmetries and these other members of the causal family. Among the latter, conservation laws are the most closely linked to symmetries, and will therefore receive special attention. That the view presented here departs significantly from the prevailing, noncausal, portrayal of symmetries highlights the implications of the broad conception of causation set forth in this book.

Symmetry as a Causal Constraint

A symmetry of a physical or mathematical object designates the object's invariance under a certain kind of transformation. Symmetry and invariance are thus two sides of the same coin. Symmetry and equivalence are likewise closely related: when we have symmetry, the original and the transformed states are equivalent in some specified sense, as, for instance, when the two states are equally probable, or when both are solutions to the same equation. Hence symmetries, like equivalence relations, generate partitions into equivalence classes. Upon the development of group theory in the nineteenth century, it became clear that

3. Einstein and Grossmann (1913). For analysis of the hole argument, see, e.g., Stachel [1980] (1989) and Norton (2015) and references therein.

symmetries form groups that can serve to characterize them.[4] It was later proved that the converse is also the case: every group is a symmetry group of a certain graph (Frucht 1939). In physics, the symmetries of *laws*, or of the equations expressing laws, are of particular interest. Symmetries are manifestations of what remains invariant in the process described by the law or equation in question. The connection between symmetry and equivalence makes it clear that symmetries also indicate which properties and parameters are, from the physical point of view, *irrelevant*. To give the simplest example, if spatial or temporal translation is a symmetry of the (Euler-Lagrange) equations of motion, this means that absolute position in space or time (as opposed to relative position) cannot make a physical difference to any evolution governed by these equations.

Even at this preliminary stage, we have some insight into the relation between symmetries and causation in the broad sense. For although symmetries are expressed as mathematical properties of mathematical objects—the equations of motion, say—under their physical interpretation, they express constraints on change, and distinguish properties that are deemed to make a physical difference from those that are deemed to make no difference. Note that there is no purely mathematical reason why absolute position in space or time should make no physical difference. The physical relevance or irrelevance of such a parameter is a matter of physics—a matter of the constraints that physical processes must satisfy. As I emphasized in chapter 1, the definitions of causation commonly put forward in the philosophical literature are geared to explaining what happens, but have little to say regarding omissions. On the broader conception proposed here, causation encompasses all constraints on change, and it can therefore be invoked to explain not only that which occurs, but also that which does not, to explain both that which is relevant to physical change, and that which is not.

Wigner, who was awarded the Nobel Prize for applying symmetry and group theory to quantum mechanics, wrote several papers on the

4. The group properties of closure, identity, and inverse follow directly from the properties of the equivalence relation—its reflexivity, symmetry, and transitivity. Associativity follows from the properties of transformations as functions.

significance of symmetries in physics. In his view, symmetry principles are indispensable for the discovery of laws:

> The laws of nature could not exist without principles of invariance. . . . If the correlations between events changed from day to day, and would be different for different points of space, it would be impossible to discover them. (Wigner 1967, 29)

Wigner further points out that the relationship between symmetry principles and the laws of nature is analogous to that between those laws and the events they apply to. Just as the laws give unity and structure to the multitude of events, so symmetry principles give unity and structure to the multitude of laws. Bringing this analogy to bear on the question of the causal status of symmetries (a problem Wigner did not address), we can say that to the extent that laws are expressions of causal connections—and it is generally agreed that they are—symmetries are as well. Symmetry principles, however, are not merely analogous to physical laws, but also serve as constraints on the form of laws. "A law of nature can be accepted as valid only if the correlations which it postulates are consistent with the accepted invariance principles" (Wigner 1967, 46). Theories constructed along these lines—"principle theories," as Einstein (1919) referred to them—start off by laying down general constraints, from which more detailed laws are then derived. The principle of relativity, which motivated Einstein's formulation of the special theory of relativity, is a paradigmatic example. Indeed, Einstein's use of this symmetry principle in 1905 is often considered a turning point in the history of symmetry in physics; from that moment on, symmetries, rather than being discovered by drawing on established theories, have served as guidelines for the construction of new theories.

Recall the legal analogy introduced in chapter 1.[5] On the freedom-inducing model, laws exclude certain forms of conduct but otherwise leave their addressees free to act in any way not ruled out by the laws. By contrast, the freedom-excluding model recognizes only duties and prohibitions; it prohibits any conduct that is not obligatory. In general,

5. But note the difference, mentioned in chapter 1, note 18, between the normative and the descriptive.

legal systems incorporate a combination of these extremes: they include prohibitions and obligations, but (fortunately) also allow a substantial amount of freedom by permitting the innumerable actions that are neither prohibited nor mandatory. The system should care about such liberties only in the sense that it should protect the freedom it has granted, but otherwise be indifferent to whether that freedom is exercised or not. The freedom-excluding model seems apt for deterministic systems. A system governed by a deterministic theory can only evolve along a single trajectory—namely, that dictated by its laws and initial conditions; all other trajectories are excluded.[6] Symmetry principles, on the other hand, fit the freedom-inducing model. Rather than distinguishing what is excluded from what is bound to happen, these principles distinguish what is excluded from what is *possible*. In other words, although they place restrictions on what is possible, they do not usually determine a single trajectory. Indeed, the very formulation of symmetry principles entails freedoms such as the liberty to rotate the system, or permute its particles, without affecting its dynamics. By identifying such legitimate transformations, symmetry principles also delineate the realm of physical significance. The changes they permit do not make a difference to what is significant from the physical point of view. Separating significant from insignificant changes is of the utmost importance to the physicist. The example of uniform motion, which was considered significant change in the Aristotelian framework but not in Newtonian mechanics, suffices to remind us that identifying parameters that are physically relevant is neither a priori nor trivial. Shifts in the distribution of physical significance, like shifts in the distribution of legal liberties, can engender revolution.

Although individual symmetry principles typically allow freedom, we may still wonder whether the combination of all known symmetries adds up to a freedom-excluding system. The example of QM, where the various symmetries imposed may still fail to determine a unique outcome, suggests that this not the case. Nevertheless, the prospect of freedom's being excluded continues to engage the imagination of physicists,

6. As has been pointed out, there are a number of caveats, such as the restriction to closed systems, that may stand in the way of a fully deterministic world.

who seek to eliminate the accidental and merely contingent. We will see that new constraints have, in fact, been adopted precisely for the sake of restricting the freedom granted by earlier constraints.

Symmetries were introduced into physics long before they were given a technical, let alone group-theoretical, treatment, for example, Archimedes's law of the lever. For Leibniz, considerations of this kind, and Archimedes's law in particular, attested to the validity of the principle of sufficient reason, which he took to be both necessary and sufficient to account for all natural phenomena (in contrast to mathematics, which could be derived, Leibniz averred, from the principle of noncontradiction alone).[7] In addition to Archimedes's law, Leibniz cited another example of the principle of sufficient reason's applicability to science: Fermat's principle, according to which light moves along the trajectory that takes the least time. He took both examples to explain not only facts, but also the *laws* governing these facts. In comparison with "ordinary" laws, the laws satisfying the principle of sufficient reason had, Leibniz felt, a distinctive elegance that reflected divine wisdom. It is quite remarkable that the two examples adduced by Leibniz represent symmetry and variation principles. Not only are these the two kinds of general principles that are still most cherished by contemporary physicists, but they are also interconnected (see chapter 6). In distinguishing laws that explain facts from laws that explain laws, Leibniz introduced the very hierarchy later embraced by Wigner. Note, however, that both higher and lower levels express physical constraints, rather than purely mathematical ones.

I have motivated the inclusion of symmetries in the family of causal constraints by giving a general description of their role in constraining physical change and physical possibility. It may be helpful, however, to adduce a concrete example of a symmetry principle that functions in this way and is directly involved in what we would ordinarily think of as causal explanation. Pauli's exclusion principle is a symmetry principle that, though at first glance far removed from any causal consideration, turns out to function as a causal constraint and to be derived from a causal constraint. It therefore merits examination.

7. For more on the principle of sufficient reason, see chapter 6.

Pauli put forward the exclusion principle in late 1924 to explain several troubling discrepancies between the Bohr-Sommerfeld model of the atom and the observed spectra of hydrogen and several other elements.[8] The observed spectra, especially under certain specific conditions, such as the presence of an external magnetic field, were known to display more splitting than allowed by the Bohr-Sommerfeld model. Pauli suggested that the electron has a twofold mode of existence (*Zweideutlichkeit*), a novel characteristic that has no classical analogue. He therefore thought it necessary to endow the electron with a new degree of freedom, represented by a fourth quantum number—in addition to the three that characterized electrons within the atom in the Bohr-Sommerfeld model. This new degree of freedom explained the mysterious splitting of energy levels that had been previously unexplained. (Inspired by Pauli, Uhlenbeck and Goudsmit associated the new degree of freedom with the electron's spin.) Pauli's tour de force (1925) was his formulation of the exclusion principle, according to which no two electrons can occupy the same quantum state.[9]

Although its empirical adequacy was immediately recognized, at this point the exclusion principle had no theoretical underpinning. The brief span between the principle's discovery in 1924 and (the first steps toward) its theoretical derivation in 1926 were the formative years of quantum mechanics. In contrast to the theoretical disarray that had characterized the "old" quantum theory, by 1927 the formalisms of nonrelativistic and relativistic quantum mechanics (due to Heisenberg, Schrödinger, and Dirac) had been put in place. There was also a growing understanding of the properties of elementary particles, including their characterization in terms of the statistics of their behavior. Such statistical differences, it soon became clear, reflect differences in the particles' identity and individuation. That is, they pertain to the question

8. On the Pauli principle's discovery, see Massimi (2005) and Blum (2014). The discrepancies in question had troubled the physics community for a long time, and gave rise to numerous attempts by Bohr, Heisenberg, and others to reconcile theory and experiment. A detailed description of these efforts is given in the cited publications.

9. Some of Pauli's ideas, in particular the *Zweideutlichkeit* conjecture, had already been conveyed to colleagues in correspondence two years prior to the 1925 publication; see Massimi (2005, chap. 2).

of whether (in some specific physical system) a permutation of two particles of a certain kind yields a new state, or a state indistinguishable from the original state. Such indistinguishability would further suggest an equivalence between the two states—namely, the symmetry of the combined state under permutations of its constituent particles. The link between permutation symmetry and the exclusion principle was recognized by both Fermi and Dirac, independently, in 1926.[10] Dirac (1926) demonstrates that for an atom with two electrons, the requirement that indistinguishable states be counted as one state allows for just two possibilities: the eigenfunction representing the combined system can be either symmetric or antisymmetric.[11] To decide between these possibilities, he noted that the symmetric solution puts no limit on the number of electrons in the same orbit. By contrast, an antisymmetric function vanishes when the two electrons occupy the same state, from which it follows that such double occupancy cannot represent a stationary state. The antisymmetric solution is therefore the only one compatible with the Pauli exclusion principle.

With these developments, it became clear that the significance of the exclusion principle was far deeper than had been initially recognized. In addition to reconciling the aforementioned discrepancies and accounting for the electron's spin, the principle came to be seen as a key to understanding the structure of matter in general. A beautiful application was worked out in 1930 by Subrahmanyan Chandrasekhar, who used the exclusion principle to study the evolution of stars whose nuclear fuel has been exhausted. Could such "dead" stars be stable, or were they at some point bound to collapse under their own gravity? Chandrasekhar (1930a; 1931b) argued that in the case of white dwarfs, the dominant factor in balancing the inward gravitational pressure is not thermal pressure but rather the pressure generated by a gas of electrons

10. What we now call the Bose-Einstein statistics for a photon gas had already been suggested by Bose and then (independently) by Einstein in 1924. Heisenberg's paper on permutation symmetry (Heisenberg 1926) also played an important role in the elucidation of Pauli's principle.

11. The eigenfunction, when antisymmetric, changes its sign under the permutation but remains unchanged when symmetric.

obeying Pauli's exclusion principle. The principle dictates that electrons brought closer by gravity are forced into higher energy levels than they would have occupied had they been free to cram together at the lowest energy level. The electrons thus develop higher speed, and as a result, higher pressure (known as electron degeneracy pressure).[12] On the basis of relativistic and quantum mechanical considerations, Chandrasekhar calculated how the balance between the gravitational pressure and the electronic pressure could be maintained. He showed that beyond approximately 1.4 solar masses, gravity overtakes the electron degeneracy pressure and the star collapses, turning into what would later be called a black hole. For our purposes, the salient point is that the exclusion principle, by accounting for these interactions and processes, plays an essential role in this calculation. The force that *counterbalances* gravity is patently as significant a causal factor as gravity itself.[13]

Pauli's subsequent work on the theoretical justification for the exclusion principle, now involving quantum field theory, culminated in his paper "The Connection between Spin and Statistics" (Pauli 1940). In the paper, Pauli extends the principle from electrons to fermions in general, and derives it from the permutation-symmetry properties of (the mathematical entities representing) the particles. He proves that bosons, particles with integer spin, must be represented by symmetric state functions (symmetric tensors), whereas fermions, particles with half-integer spin, must be represented by antisymmetric state functions. The exclusion principle, he shows, applies only to fermions. It might be thought that at this higher level of formality and generality, the principle would have no manifest connection to causation. In fact, however, a crucial step in the derivation is based on (relativistic)

12. In fermionic systems, which obey Pauli's principle, the electron degeneracy pressure is present even at absolute zero.

13. The argument is still considered valid and well confirmed by observation. The possibility of a violation of the Chandrasekhar limit under certain conditions was recently raised, but is still an open question, and in any case, would not weaken the argument regarding the causal role of the exclusion principle. Eddington's war against Chandrasekhar, and dismissal of this amazing prediction of black holes (as we now call them), is a sad episode in the history of science.

locality—a manifestly causal premise! Pauli assumes that the charge density at a given point must be independent of the charge density at any point located at a space-like distance from the original point. This causal constraint, sometimes referred to as *microcausality*, is an integral part of the justification for the exclusion principle.[14] Both in terms of its derivation, then, and in terms of its application and explanatory import, the exclusion principle is directly tied to causal considerations.

This account of Pauli's exclusion principle differs from those of Railton (1978), Salmon (1989), and Lange (2017), all of which characterize the exclusion principle as a kind of structural principle that provides a noncausal explanation.[15] At first glance, the noncausal interpretation seems plausible. Pauli's principle does indeed highlight the remarkable effectiveness of mathematics in physics. Its formulation in terms of symmetry properties of abstract mathematical objects such as Ψ functions might lead us to believe that abstract mathematical structures and relations shape the world, or to put it in less Pythagorean language, that they suffice to explain the world. I have emphasized, however, that although the exclusion principle has an abstract mathematical appearance, it does not spring from purely mathematical considerations. Mathematics cannot, by itself, account for the existence of particles that are indistinguishable from one another, act in conformity with Fermi-Dirac statistics, and are excluded from occupying the same state. By the same token, there is no purely mathematical explanation for microcausality, the relativistic constraint on the transmission of information. I take these reflections to indicate that the exclusion principle should not be

14. See Haag (1993, chap. 4) on the connection between particle statistics and the "causal assumption" (his term) that very distant particles are asymptotically independent. "Causal" is being used here in the relativistic sense of locality, i.e., as a restriction on the interaction speed. Temporal asymmetry is integral to this causal constraint.

15. Railton (1978), Salmon (1989, 159–66), Lange (2017, 183). Salmon mentions the gas law and Lange mentions conservation laws and, somewhat surprisingly, even Newton's second law, as exemplifying this noncausal category. Regarding Newton's second law, Lange argues that it applies to forces in general rather than any specific force. But why should the generality of a law conflict with its causal status? Newton's second law—the connection between a force and the resulting acceleration—should be considered a causal law even on the traditional understanding of causation as a relation between individual events or properties.

construed as a purely mathematical symmetry. Not only is the principle *based on* the causal consideration of locality, it has also, as we saw apropos the collapsing star, been very successfully *used as* a causal constraint. Taken as a full-fledged causal constraint, it is an excellent illustration of the causal function of symmetries in physics.

Conservation Laws

Symmetries are transformations that keep certain parameters (properties, equations, and so on) invariant, that is, the parameters they refer to are conserved under these transformations. It is to be expected, therefore, that the identification of conserved quantities is inseparable from the identification of fundamental symmetries in the laws of nature. Symmetries single out "privileged" operations, conservation laws single out "privileged" quantities or properties that correspond to these operations. Yet the specific connections between a particular symmetry and the invariance it entails are far from obvious. For instance, the isotropy of space (the indistinguishability of its directions) is intuitive enough, but the conservation of angular momentum based on that symmetry, and indeed, the concept of angular momentum, are far less intuitive. The connection between symmetries and conservation laws emerged gradually from reformulations of Newtonian mechanics by Euler, Lagrange, Hamilton, Jacobi, and others.[16] The notion of symmetry itself, however, only came to the fore with the development of group theory in the nineteenth century. So while in current formulations of physical theories, conservation laws are generally derived from symmetries, historically, conservation laws typically came first. Moreover, the assumption that some fundamental continuity must underlie change goes back even further.

Observing the world around us, we notice various types of change and persistence. A rock on the beach seems to have been there forever, whereas a burning log turns into ash and smoke, the smoke ultimately vanishing into the air; rubbing our hands creates heat, but the heat dis-

16. For a historical account of these developments, see Lanczos (1949).

sipates quickly; a pendulum keeps swinging for a long time, but eventually stops, and so on. To the Greeks, who were puzzled by the variety of change, two extreme possibilities suggested themselves: the Parmenidean view, according to which change is illusory, and the Heraclitean view, on which change is ubiquitous and persistence illusory. Over the next two millennia various attempts were made to reach a compromise between these extremes: theories that endeavored to identify a few constant elements in an otherwise changing world. The identification of such constants eventually led to elucidation of the concepts of matter, motion, force, momentum, energy, and so on, none of which were directly observable, let alone self-evident. A perceptive Quinean observation comes to mind: "Our coming to understand what the objects are *is* for the most part our mastery of what the theory says about them. We do not learn first what to talk about and then what to say about it" (1960, 16). Discovering the various conservation laws of classical mechanics, and gaining an understanding of the physical concepts to which those laws apply, were therefore entwined in a process of great complexity. To give but one example, consider the long path from the earliest intuitions about the conservation of matter to the laws of conservation of mass, energy, mass-energy, baryon and lepton number . . . (in light of dark matter and dark energy, the end of the path is not yet in sight). Let me look somewhat more closely at this example—a delusively simple conservation law that repeatedly defied attempts to fully capture it.

Rudimentary ideas as to matter's having a constant quantity were already part of both the "four elements" theory proposed by Empedocles and the atomic theory defended by Leucippus, Democritus, and Epicurus. The former involved the idea that the elements were neither created nor annihilated, but gave no precise definition of the quantity of matter. The latter was more developed in this respect. Atoms were taken to be eternal, indestructible, and indivisible; their number and weights were taken to be fixed. The principle *nil posse creari de nihilo* (as Lucretius renders it in *De Rerum Natura*) became a fundamental constraint on change. A related tenet, formulated in the classical era and reiterated during the scientific revolution, is the equality of cause and effect—*causa aequat effectum*. Were there "more" in the effect than there

had been in the cause, this would amount to creation ex nihilo. But more of what? At this stage, there was no deep understanding of this tenet, and certainly no way of quantifying it. The view that weight was the property characteristic of matter was contentious. Aristotle and a long line of his followers viewed weight as just a secondary property of matter, and unlike the atomists, the Aristotelians did not take all elements to be heavy.[17]

As was the case with other concepts of Aristotelian physics, the concept of matter was significantly revised during the scientific revolution. Newton's concept of mass involved a clear distinction between matter's quantity and its weight. Moreover, Newton's theory propounded two distinct characteristics of matter and thus two distinct concepts of mass: inertial mass, measured by the ratio between the force acting on a body and that body's resultant acceleration, and gravitational mass, which was responsible for a body's exerting an attractive force on every other mass in the universe in accordance with the law of gravitation. Newton noted the conceptual difference between the two but declared them mathematically identical (up to a constant). A body's weight is its mass multiplied by its gravitational acceleration; its weight therefore varies with this acceleration. Once this relation is taken into account, weight, being proportional to mass, can provide a measure of the quantity of matter.

It is obvious from this schematic description that changes in the concept of matter were inseparable from refinements of the concepts of force and motion (and related concepts such as inertia, velocity, acceleration, momentum). In mechanics, the conservation of mass, perhaps because it was taken for granted, was not formulated as a fundamental principle. Conservation laws related to motion and force—the conservation of quantity of motion (a precursor of linear momentum) and vis viva (which is proportional to what we call kinetic energy)—were, however, avidly debated by Descartes, Huygens, Newton, and Leibniz. For example, whereas Descartes's law of the conservation of quantity

17. The idea that a chemical reaction could change a metal's weight was a long-standing tenet of alchemy, current as late as the eighteenth century.

of motion took that quantity to be scalar, Huygens noted the vectorial nature of the conserved quantity. Neither one had the clear concept of mass that Newton was soon to develop. Leibniz's conservation of vis viva also came under fire from several directions. Huygens, who had already discovered a similar law, claimed it held only for elastic collisions; Cartesians claimed that it was nothing more than Descartes's conservation of quantity of motion law; and Newton ignored it. The ancient principle that cause must equal effect, and the argument that violation of this principle amounts to creation or annihilation, reverberate through these debates, though often in theological rather than metaphysical language. Both conservation laws (conservation of momentum and conservation of kinetic energy) were finally embedded in Newton's system and shown to follow from his laws of motion. For instance, it is clear from Newton's second law that in the absence of external forces, linear momentum is conserved.[18] The discovery of the conservation of angular momentum, which likewise follows from Newton's system, took almost another century; it was formulated and proved by Leonard Euler and Daniel Bernoulli.

In chemistry, in contrast to mechanics, the question of whether mass is conserved in all chemical reactions was a major issue, and was still undecided in the mid-seventeenth century. The budding science of chemistry had, however, already made several contributions of its own to the understanding of matter: gases had come to be seen as matter; atomistic theories were applied in chemical research; and the concepts of element, compound, and mixture were formed (though no method of distinguishing them in practice was yet available). In 1785, Lavoisier, whose meticulous experiments had convinced him that mass is neither created nor destroyed in chemical reactions, announced the principle that mass is conserved in a closed system. Lavoisier's principle became an essential tool for analyzing chemical reactions.

18. Although linear momentum is defined as the product of mass and velocity (the time derivative of position), it is a more fundamental quantity than either mass or velocity. Indeed, linear momentum turned out to be conserved even under the relativistic regime, where Newtonian mass is not conserved.

Meanwhile, other forms of matter—electricity and heat—were being conjectured. Benjamin Franklin conceived of electricity as consisting of atoms of an electric substance, atoms sufficiently small to penetrate ordinary matter. When these electric atoms were uniformly distributed, he contended, they could not be detected, whereas uneven distributions gave rise to electrical phenomena. Their accumulation manifested itself in what Franklin called "positive" electricity, and their dearth in "negative" electricity. Drawing an analogy with ordinary atoms, he surmised that electric atoms are neither created nor destroyed, from which conservation of the quantity of electricity followed as a simple corollary. Franklin's monistic model of electricity was subsequently superseded by Coulomb's dualistic model. Initially, the dualistic model seemed to conflict with the principle of conservation of mass. How could two different substances neutralize each other, as positive and negative electric charges should, according to the model, without violating this principle? In the face of doubts about the nature of electricity, Faraday formulated the conservation of electric charge in a manner independent of any specific model. His law permitted the creation of charge as long as only *pairs* of equally strong positive and negative charges were created. It still followed, though, that in a closed system the net charge remains constant even if additional electric charges are created within the system, say by friction of its constituent parts. Although the law of conservation of electric charge was inspired by the law of conservation of mass, the former actually proved to be more basic, surviving as a scientific law long after its precursor had to be modified.[19]

The nature of heat was another formidable problem. Here too there were initially two rival models, with no definitive experiment to decide between them: the caloric theory, on which heat was a distinct substance (termed "caloric" by proponents of this view), and the motion theory, on which heat was the manifestation of internal movements of the components of ordinary matter. Again, the field progressed despite

19. At present, only the combination of charge, parity, and time reversal is considered to be a fundamental symmetry.

this ambiguity. Sadi Carnot, for example, intentionally formulated the principle that bears his name in terms that made it compatible with both theories.[20] Since heat engines turn heat into mechanical work, the study of heat engines focused attention more generally on conversions of (what would later be called) different forms of energy. In addition to the already-established transformations of kinetic to potential energy, and vice versa, in mechanics, research now turned to the transformation of heat into mechanical work, and vice versa. The production of heat by electricity was also recognized. Were heat a substance that flowed from one place to another, as the caloric theory claimed, a conservation law analogous to the conservation of mass would apply. But from the perspective of the motion theory of heat, which eventually superseded the caloric theory, the picture was more complicated. Moreover, research into the transformations of heat appeared to yield two conflicting conclusions. On the one hand, Joule established the equivalence between a specific amount of heat and a specific amount of mechanical work produced by that heat. On the other hand, Carnot showed that there is a limit to the efficiency of even ideal heat engines, so heat can never be completely converted into mechanical work. The conflict was noted by Lord Kelvin and reconciled by Rudolf Clausius: Carnot was right to claim that heat could not be completely converted into work, but the amount that *is* converted obeys Joule's equivalence relation. To use later terminology, Joule's discovery amounts to an early version of the first law of thermodynamics—the law of conservation of energy—whereas Carnot's is the earliest version of the second law of thermodynamics. By the end of the nineteenth century, both these laws were in place.[21]

20. Carnot's principle, published in 1824, is the first version of the second law of thermodynamics. It states that the production of work in heat engines is invariably accompanied by the flow of heat from a reservoir of higher temperature to one of lower temperature. The flow image fits the caloric theory better than the motion theory, but it appears that Carnot, at least toward the end of his life, subscribed to the motion theory.

21. At this point in history, neither heat nor electricity was taken to be a distinct form of matter; there was only one kind of matter, characterized by mass, and governed by the law of conservation of mass. The other conservation laws applied to specific properties of matter: linear momentum, angular momentum, electric charge, energy. Subsequently, there were a

The concept of energy gained special importance, becoming the cornerstone not only of thermodynamics, but also of Hamilton's reformulation of mechanics, the crown of classical physics. For a short while, mass and energy were considered distinct concepts, demarcated by two independent conservation laws. But not for long. Right after publishing his special theory of relativity, Einstein came to the revolutionary conclusion that "the mass of a body is a measure of its energy content" (1905, 174) and formulated the unified principle of the conservation of mass-energy. The long-standing "no creation, no annihilation" principle, which—given the possibility of matter's being transformed into massless radiation—had appeared to be in jeopardy, thus regained its footing. What had seemed to be a violation of the "no creation, no annihilation" principle was resolved by extending both the conservation laws and the concepts to which they applied.

The concept of matter continued to evolve. For Newton, as we saw, inertia and gravity were distinct properties of matter that happen to be mathematically identical (up to a constant), a mere coincidence from the scientific point of view, or (for Newton) a kind gesture by a beneficent God. Einstein reframed this coincidence as the principle of equivalence, the fundamental principle of the general theory of relativity (GTR), achieving the unification of inertia and gravity. Einstein's conviction that the mathematical identity of matter's two properties must reflect a profound connection between the two is a striking example of the aforementioned predilection, on the part of physicists, for excising the merely contingent from the scientific picture of the world.

Quantum mechanics and quantum field theory revolutionized the concept of matter even further, but we can stop here. The developments surveyed thus far suffice to illustrate the difficulty of turning the early intuitions about the conservation of matter into scientific principles, and the many changes the concept of matter would undergo in the process. At the same time, these developments also attest to the conviction

---

number of attempts to reduce matter to one of these characteristic properties, including Lorentz's electromagnetic theory of matter, and the attempt by Ostwald and his followers to reduce physics in its entirety to "energetics."

that motivated many of them: the belief that underlying all change are constancies that delimit its scope. Conservation laws, which sit atop the hierarchy of physical laws, have definitely vindicated this belief. Moreover, mature conservation laws, like their early prototypes, are generally considered manifestations of the causal order. Bohr and Meyerson provide clear examples of the causal understanding of conservation laws. As mentioned in chapter 1, when explaining what he means by "complementarity," Bohr uses the term "causality" to refer to the conservation of energy and momentum. He claims that a system's causal and spatiotemporal descriptions are complementary, that is, they cannot be applied together. Consider, for example, position and momentum, which, by Heisenberg's uncertainty principle, cannot both have precise values at a particular time. Bohr's argument is that a spatiotemporal description that follows the exact trajectory of a particle in spacetime would not be able to ascribe a definite linear momentum to that particle, let alone establish the conservation of its linear momentum. Hence the particle's spatiotemporal description excludes its causal description. This use of the term "causal description" seemed so natural to Bohr that he did not stop to explain it. Emil Meyerson also linked conservation and causation. As its title indicates, his *Identity and Reality* explores the profound significance of quantities that remain invariant through change (or apparent change). "The external world," he says, "appears to us as infinitely changing. . . . Yet the principle of causality postulates the contrary. . . . change is only apparent; it must necessarily disclose an identity which alone is real" ([1908] 1930, 92). This conception of causation, readers will notice, is considerably different from the Humean conception of causation as regularity.

The Connection between Symmetry Principles and Conservation Laws

As mentioned, the close connection between specific symmetries and specific conservation laws, though implicit in the Newtonian system, was not articulated by Newton, but by mathematicians such as Euler and Lagrange, who reformulated Newtonian mechanics during the

eighteenth and nineteenth centuries. The connection is particularly clear in the case of Hamilton's formalism. Roughly, the basic mathematical function in this formalism (later called the Hamiltonian) represents a system's total energy, and contains crucial information about the system's evolution in time. If the Hamiltonian is not an explicit function of time, it is invariant throughout this evolution. Furthermore, for any coordinate not appearing explicitly in the Hamiltonian, the momentum conjugate to this coordinate is conserved.[22] Now, if the system is symmetric under a certain transformation, this symmetry must find its expression in the Hamiltonian. For example, the assumption of the homogeneity of space implies a symmetry—the mere spatial translation of a free system should not change the physics of the situation. Hence the coordinates for the system's center of mass should not appear in the Hamiltonian. The momentum conjugate to this coordinate—the system's total linear momentum—will therefore be conserved. Similar considerations apply to the rotational symmetry, assuming the isotropy of space. In this case it is the angle of rotation that should not appear explicitly in the Hamiltonian, and the momentum conjugate to this angle, which is conserved throughout the system's motion, is the system's angular momentum.

A more thoroughgoing connection between symmetries and conserved quantities was only established in the twentieth century, when the converging interests of experts in mathematical physics and group theory, especially those based in Göttingen, facilitated a deeper understanding of the symmetry–invariance nexus. GTR triggered this research by raising questions about two pivotal issues: the significance of what Einstein saw as GTR's fundamental symmetry—namely, general

22. Hamilton's formalism is based on that of Lagrange, which is formulated in terms of the scalars of potential and kinetic energy rather than the Newtonian vectors of force and momentum. The Hamiltonian is defined as $\sum \dot{q}_i p_i - L$, where q_i is the i-th coordinate (of parameter q, say position); $\dot{q}_i$ is the time derivative of q_i; p_i is the momentum conjugate to the i-th coordinate; and L is the Lagrangian, defined as the difference between the potential and kinetic energies $T - V$. The rough formulation given here ignores certain qualifications, such as the Lagrangian's independence of the velocities. The Hamiltonian's invariance and the fact that it represents the total energy are also not equivalent; see Goldstein (1973, chap. 7, esp. 221). For more on the differences between the Lagrangian and Hamiltonian formalisms, see chapter 6 herein.

covariance, and the status of the law of conservation of mass-energy within GTR. These questions motivated Emmy Nöther's celebrated 1918 paper, which was a milestone in the elucidation of symmetries and their relation to conservation laws.[23] The general results Nöther reached in this paper strengthen the conclusion that insofar as conservation laws are causal laws, so are the symmetries from which they are derived.

Nöther's paper investigates continuous symmetries of the action, that is, continuous groups of transformations that leave the action integral invariant.[24] She discusses two types of such symmetries, represented by finite- and infinite-dimensional Lie groups. According to Nöther's first theorem (for a system governed by the Euler-Lagrange equations of motion), a continuous symmetry of the action, represented by a finite n-parameter Lie group, is correlated with n conserved quantities. This theorem entailed the conservation laws of classical mechanics—namely, the conservation of kinetic energy, the conservation of linear and angular momentum, and the uniform motion of the center of mass under the influence of internal forces (all well established by that time). According to Nöther's second theorem, the invariance of the action under symmetries represented by an infinite-dimensional Lie group (characterized by m functions, rather than m parameters) leads to m identities among the Euler-Lagrange equations of motion. In other words, the symmetries of the action add further *constraints* to those imposed by the equations of motion. Note that both types of constraints (those expressed by the equations and those expressed by symmetries of the action) are causal constraints in the sense in which I use the term. Nöther also showed (a corollary sometimes considered her third theorem) that when the infinite symmetry group of the action contains a finite subgroup of symmetries, the conservation laws derived from the first theorem follow from the identities derived from the second. For the electromagnetic field, for example, the

23. For a detailed analysis, see Brading and Brown (2003).

24. The action is the integral of the Lagrangian over time. As mentioned in note 22 in this chapter, the Lagrangian is the difference between a system's potential and kinetic energies.

identities derivable from the symmetry of the action are Maxwell's equations, and the conservation law that follows from them is the conservation of electric charge.

The Variety of Symmetries

It is customary to distinguish between active and passive symmetries. Active symmetries pertain to physical changes, such as the translation of a system to a different region in space, whereas passive symmetries apply to modes of description. The passive symmetry corresponding to the actual translation of a system to another point in space (which constitutes an active symmetry) would be the complementary translation of the origin of the coordinate system used to describe the system. At first glance, it might seem that only active symmetries contain information about physical reality and its causal structure, and that passive symmetries—mere artifacts generated by descriptions—are physically uninformative. In the translation example, however, the active and passive symmetries are equivalent, and physicists thus tend to see them as interchangeable. If the active symmetry under translation presupposes the homogeneity of space, so does its passive counterpart; if the former leaves the equations of motion invariant, so does the latter, and for exactly the same reason. Neither symmetry affects physically meaningful parameters and their interrelations. Thus, if the active symmetry is related to the conservation of linear momentum, so is the passive symmetry. The example can be generalized, suggesting that passive symmetries that tell us which transformations of the descriptive apparatus are legitimate can be just as informative about the natural world as active symmetries that involve transformations of real physical systems.

Another important distinction, introduced by Wigner (1967), is that between global and local symmetries, also referred to, respectively, as geometric versus dynamic symmetries. Whereas global (geometric) symmetries are symmetries of space and time and therefore apply to all physical interactions, local (dynamic) symmetries characterize specific interactions and forces. The distinction between these types of

symmetries corresponds to the distinction between the symmetries satisfying Nöther's first and second theorems: global (geometric) symmetries are described by Nöther's first theorem, local (dynamic) symmetries by her second theorem. Examples of global (geometric) symmetries, which are represented by finite-dimensional Lie groups, are isometry (translation invariance) and isotropy (rotation invariance). Local (dynamic) symmetries are expressed by functions (rather than parameters) that generally vary from one point to another, and are represented by infinite-dimensional Lie groups. Examples of local (dynamic) symmetries are the symmetries of electromagnetic, strong, and weak interactions, and the theories that unify them, such as the Yang-Mills theory and the standard model of particle physics. These local symmetries are the focus of gauge theories.

In general, gauge theories involve transformations that can be viewed as the rescaling of a physical parameter; gauge symmetries, in turn, are the symmetries under this operation.[25] An example of such a theory (though not under that name) is Hermann Weyl's attempt to extend GTR by generalizing the Riemannian geometry on which it is based (Weyl [1918] 1952). Generalization was called for, Weyl maintained, in order to weaken Riemann's fundamental assumption regarding the integrability of length. According to Riemann, the parallel transport of a vector along a closed path, while in general resulting in a change in the vector's *direction*, nonetheless retains its *length*. Weyl saw this assumption as departing from the local nature of Riemannian geometry by retaining a kind of congruence at a distance, analogous to the notorious action at a distance of Newtonian mechanics. To relax the assumption, Weyl introduced a new vector field that indicates the variation (rescaling) of the unit of length throughout the manifold and complements the affine structure on the manifold, which indicates the variation in the vector's direction. This new field turned out to be formally identical with the electromagnetic field. Weyl's theory thus constituted the first

25. More technically, a gauge theory is characterized by the invariance of its Lagrangian under a group of local transformations. These local transformations represent the aforementioned rescaling.

unified field theory encompassing two dynamic structures, the affine connection representing the inertio-gravitational field of GTR and the new gauge structure representing the electromagnetic field. (Recall that gravity and electromagnetism were the only kinds of interactions known at the time.) Weyl's approach was revived in the 1960s in the context of elementary particles theory, where gauge invariance was found to express the phase invariance of the quantum-mechanical wave function.

The electromagnetic theory also provides an example of gauge *freedom,* manifested in the indeterminacy of the potentials corresponding to the fields.[26] In Maxwell's equations, the physical (and measurable) information is expressed by the electric and magnetic fields. The equations represent the relations between these fields and two potentials: the scalar potential of the electric field and the vector potential of the magnetic field. These potentials, however, do not express measurable quantities, and can be changed (rescaled) in specific ways without affecting the strength of the electric and magnetic fields.

Michael Redhead (2003) provides a particularly lucid exposition of the notion of gauge freedom. Briefly, he invites us to abstract from the common understanding of gauge transformations as rescaling, and think of gauges more generally, in terms of possible relations between a physical structure and its mathematical representation. Ideally, we would want the physical and mathematical structures to be isomorphic, so that any physically significant information, and only such information, will be mathematically represented. In particular, when the physical and mathematical structures are isomorphic, the mathematical structure should reflect the deterministic nature of the physical theory—the uniqueness of a solution under a particular set of initial conditions. Typically, however, the mathematical representation of a physical structure is not unique, that is, there is what Redhead calls "surplus structure" (see figure 3). In such cases, we are likely to obtain more than a single mathematical solution to the same physical problem

26. As we saw in the previous chapter, this freedom gives rise to the disparate interpretations of the Aharonov-Bohm effect.

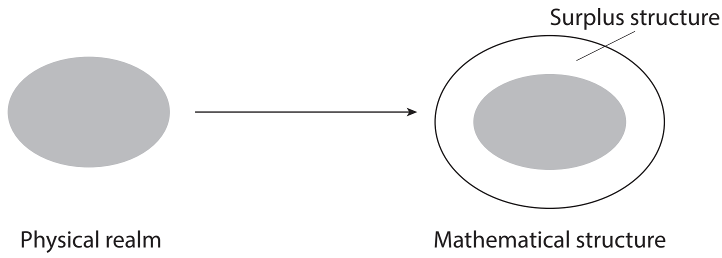


FIGURE 3. Embedding a physical realm in a nonisomorphic mathematical structure. The surplus structure is schematically represented by the outer ellipse (Redhead 2003, 128).

(the same initial conditions), a result incompatible with determinism. This one-many relation between the initial conditions and the solution is a manifestation of gauge freedom.

To correct this situation so as to restore determinism, physicists impose on the mathematical structure a "gauge symmetry" that posits the *physical equivalence* of different mathematical solutions.[27] In other words, different mathematical representations are deemed different descriptions of the same physical situation. As a result, the freedom generated by surplus structure no longer represents the indeterministic nature of the physical process, but rather the indeterminacy of its mathematical description. Observable properties of the system are gauge invariant, that is, they retain their values under the new symmetry. On this model, by imposing gauge symmetry, the physicist extracts the observable physical content of a theory, which is invariant under the imposed symmetry transformation. This strategy was first adopted by Einstein in response to his celebrated "hole argument," and has become widely used in the context of later gauge theories.[28]

27. Wallace (2003), 169–70, distinguishes between two methods of imposing such equivalence. One method involves identifying configurations related by a transformation belonging to a local symmetry group; the other involves identifying *histories* related by such a transformation, yet distinguishing between the sequences of configurations that make up those histories. The former method, he argues, is inapplicable to GTR.

28. For an analysis of the hole argument, see Stachel [1980] (1989); Earman and Norton (1987); Rickles (2008), chap. 4; and Norton (2015).

In discussing the gauge freedom of the electromagnetic interaction, Wigner introduced the aforementioned distinction between geometric symmetries, which he took to be physically meaningful, and symmetries that reflect descriptive freedom and are mere artifacts of the mathematical representation. Since gauge symmetries belong to the latter category, they are deemed to lack physical meaning.

> This invariance is, of course, an artificial one, similar to that which we could obtain by introducing into our equations the location of a ghost. The equations then must be invariant with respect to changes of the coordinate of that ghost. One does not see, in fact, what good the introduction of the coordinate of the ghost does. (Wigner 1967, 22)

In the light of similar concerns, the status of gauge symmetry is still a matter of controversy among philosophers.[29] As symmetries of the "surplus" mathematical structure, gauge symmetries appear to have neither physical significance nor (a fortiori) causal significance. Nevertheless, there is good reason to endow gauge symmetry with physical and causal significance. First, in the case of global (geometric) symmetries, I pointed out that we move freely between the passive and active perspectives—a geometric transformation of the system is equivalent to a complementary transformation of the coordinate system. My conclusion was that a constraint on what counts as a proper description of a physical system is on a par with a constraint on the system's evolution. Gauge symmetries are analogous to passive symmetries in the sense that they too are artifacts of the description, but as I just noted, by clearing away surplus structure, they distill the physical content of the theories that satisfy them. Second, and more importantly, gauge symmetry is a powerful constraint that serves to restrict—and often,

29. See Healey (2007), Martin (2003), and Morrison (2003). The focus of Healey's in-depth analysis of gauge theories is the feasibility of a realist understanding of physical entities and their properties as described by these theories. He does not address the question of causation. Although realists tend to endorse causation (and nonrealists tend to reject it), I have separated the two issues, concentrating on the role of causation as constraining change and setting aside the much-discussed issue of realism vis-à-vis theoretical concepts.

uniquely determine—the form of theories that satisfy it. Quantum electrodynamics, quantum chromodynamics, and even GTR illustrate the effectiveness of this constraint.[30] Much like other symmetries, then, gauge symmetry has become an indispensable tool for theory construction. Lastly, gauge symmetries have been shown to have concrete empirical import. The reasoning underlying the extraction of this empirical import is as follows. A gauge transformation can have peculiar apparent effects, such as making uniform motion appear accelerated, or apparently turning a particle with a fixed identity, a proton, say, into one with a fluctuating, proton-neutron identity, both changes proceeding in accordance with the continuous gauge transformation. Although these "effects" are initially purely fictitious by-products of the gauge, they can be reinterpreted as effects of a *dynamic*—also fictitious at this point—that serves to "explain" them. The next step is to take the equivalence between the passive and active perspectives seriously, and breathe life into all these fictions. This crucial step should lead us to look for evidence for the real (i.e., nonfictitious) presence of the new dynamic and its effects. Remarkably, this strategy works—Wigner's mathematical ghosts turned out to have perceptible empirical import.[31] This is not as paradoxical as it first seems. It only means that constraints on the *description* of the physical world are sometimes as informative as manifestly physical constraints. Rather than the ghost analogy, I would suggest an analogy with a glove, whose shape can tell us a great deal about the shape of a hand.

Curie's Principle

According to Curie's principle, the symmetries of the cause are satisfied by the effect—the symmetries of the cause constitute a subgroup of

30. GTR is sometimes derived as a field theory in flat spacetime satisfying gauge symmetry under Lorentz gauge transformations. On the difference between GTR and other gauge theories, see, e.g., Earman (2003), Wallace (2003), Healey (2007), and the references in note 28 in this chapter.

31. A celebrated example attesting to the empirical success of gauge theory is the discovery of W and Z bosons, communicators of the electro-weak interaction. Carlo Rubbia and Simon van der Meer were awarded the 1984 Nobel Prize for this discovery.

the symmetry group of the effect. The reverse is not true—new symmetries can appear in the effect. In other words, the effect can be more symmetric than the cause, but the cause cannot be more symmetric than the effect. Hence, any *a*symmetry manifest in the effect must be paralleled by an *a*symmetry of the cause. The principle can be formulated without explicit reference to causes; it can assert, for example, that the symmetries of equations must be symmetries of their solutions, and any *a*symmetry of the solutions must reveal itself in an *a*symmetry of the equations. Or it can be taken to rule out the possibility that an isolated system will lose some of its symmetries as it evolves.

If we think of simple examples of symmetry, such as the law of the lever, we immediately notice that the intuition underlying Curie's principle is that symmetry excludes change, whereas *a*symmetry generates change. To bring about change, there must be some difference that makes a difference, some *a*symmetry that moves the system out of the impasse generated by its balanced symmetric state. (Recall Leibniz and the principle of sufficient reason.)

At first glance, Curie's principle differs from the two most common formulations of the traditional causal principle—the universality principle, according to which every event has a cause, and the uniformity principle, according to which the same (type of) cause has the same (type of) effect. In fact, however, Curie's principle is a descendant of, and replacement for, these traditional principles: in tracing any asymmetry of the effect to an asymmetry in the cause, universality is posited, and uniformity implied. Note also that the asymmetry between cause and effect is an integral part of Curie's principle, not an additional, independent, condition.[32]

32. Given the affinity between Curie's principle and these traditional principles, it is not surprising that critics of the causal principle also seek to discredit Curie's principle. Indeed, Norton (2016) argues that the principle is too vague to be useful. Depending on how we define the "cause" and the "effect" in any particular situation, Norton contends, we can render the principle true (trivially true) or false. As mentioned, however, the principle can be couched in more precise terms than its original cause–effect formulation. Moreover, plasticity is the norm for scientific concepts. The history of the law of conservation of matter, recounted earlier in this chapter, and the transformations of the principle of least action, examined in the next, should suffice to demonstrate that as our theories evolve, concepts such as matter, mass, energy,

The causal principle, as we have seen, is no longer upheld as an article of faith by contemporary physicists. Does Curie's principle have a better track record? On the face of it, Curie's principle appears even less sustainable. We encounter numerous processes that appear to be going in the direction precluded by the principle, cases where the effect (or the solution) is *less* symmetric than the cause (or the equation). In other words, we constantly encounter the *breaking* of symmetry. Phase transitions, such as the more symmetric water in the bucket turning, when cooled, into a less symmetric block of ice, or ferromagnetic material, when cooled, acquiring magnetization in a specific direction, seem to constitute very common counterexamples to Curie's principle. The physicist's response, however, is that these are only apparent counterexamples and should not weaken our confidence in the principle. The asymmetry involved in the distinct direction of magnetization is indeed an asymmetry of one possible solution, an asymmetry absent in the original state. But if we consider, as we should, *the set of solutions in its entirety*, the symmetry reappears, for the material could have acquired an axis of magnetization in any one of an infinite number of directions. The entire set of possible solutions, then, maintains the spherical symmetry of the original nonmagnetic material.

Yet we might persevere in our doubts, wondering, why *this* particular direction? Here stability, or rather, a combination of stability and instability, comes to Curie's rescue; the slightest disturbance, the slightest deviation from the original symmetry, is sufficient to lead to the radical asymmetry we are witnessing. But once a particular direction has emerged, it is so stable that a spontaneous transition from this particular solution to any one of the other possible solutions has a negligible probability and is virtually impossible. Our causal explanations of phenomena such as phase transitions involve not only general symmetry considerations, but also system-specific stability considerations that tell us what it takes to break the symmetry in question.

action, and other fundamental scientific notions, are constantly redefined and adjusted to new applications.

As Joe Rosen points out (2008, chap. 5), Curie's principle is adduced by physicists both to set a lower bound on the symmetry of the effect and to set an upper bound on the symmetry of the cause. The latter involves positing higher symmetries, which are then broken, yielding the world we encounter. For the physicist, this application is the more exciting one, though in the view of some philosophers, it is more controversial. My own partiality to the physicist's preference is beside the point. What is significant is that Curie's principle continues to inform the search for a unified picture of the world. Philosophers such as Russell and Norton may have been right to question the validity of a universal causal principle, and all the more so, its alleged a priori validity, but in light of the utility of very general constraints on change, and in particular, Curie's principle, they may have given up on the variety of causal principles somewhat prematurely.

6

The Principle of Least Action

FROM TELEOLOGY TO CAUSALITY

THE STORY OF THE PRINCIPLE of least action takes us on a heroic journey from the humblest of origins to the pantheon of physics. This chapter, in addition to examining the least action principle's place in the causal family, uses the principle as a prism that brings to light the twists and turns in modern science's transition from teleology to causality. The journey is particularly edifying in view of the fact that the principle spurred allegations of teleology long after the general thrust of physical explanation had become causal rather than teleological.[1] Indeed, although the allegations of teleology have been compellingly refuted, an aura of mystery still clings to the principle even today, as those who have endeavored to teach the principle can attest. Moreover, the principle's reappearance in the probabilistic context of quantum mechanics provides an example of the presence of *apparent* teleology in systems that defy not only teleology but also determinism.

1. During the scientific revolution of the seventeenth century, as we will see, the contrast between teleology and causality was usually drawn in Aristotelian terms, highlighting the difference between final and efficient causes. The notion of efficient cause, however, is not invoked in contemporary science and is in any case very different from the broader notion of cause as a constraint on change, on which I focus. Nevertheless, in the framework of this book, goal-directed processes are distinguished from causal processes that only obey the constraints of determinism, locality, and so on. The contrast between teleology and causation should therefore be understood as the contrast between goal-directed processes and other types of causal process discussed in this book.

Before embarking on this fascinating journey, however, it will be helpful to mention another principle, the principle of sufficient reason, which served as the backdrop for the early deliberations about the meaning and status of the principle of least action.

In contemporary philosophy, reasons and causes are distinct explanatory categories. Ordinary language does not always respect the distinction: in English, for instance, when responding to a "why" question with an answer beginning with "because," we could be adducing either a cause or a reason. Yet philosophers such as Wittgenstein, Davidson, Putnam, Sellars, McDowell, and Brandom, who are certainly mindful of the vagueness of ordinary language on this point, repeatedly warn against conflating the two categories. Reasons are normative; they can be good or bad and can be invoked to explain goal-directed actions. By contrast, causal discourse is descriptive and nonteleological. Moreover, ascribing reasons for an action to an agent usually presupposes the agent's awareness of those reasons. No such presupposition is implicit in causal explanation—a planet or atom is unaware of the constraints on its trajectory.[2] Reasons are thus confined to the sphere of thought and action, and as such, unsuitable for scientific explanation of natural phenomena. This is also the received view among contemporary scientists. But this realization did not come easily, emerging from a jagged trek through numerous controversies from antiquity to the present. In particular, despite resolute efforts to break with the Aristotelian tradition, natural philosophy in the seventeenth and eighteenth centuries was still replete with teleological explanations framed in terms

2. That human beings actually have more freedom than atoms and moons has, of course, been contested by determinists. Einstein asserts: "If the moon, in the act of completing its eternal way round the earth, were gifted with self-consciousness, it would feel thoroughly convinced, that it would travel its way of its own accord on a strength of a resolution taken once for all. So would a Being, endowed with higher insight and more perfect intelligence, watching man and his doings, smile about the illusion of his, that he was acting according to his own free will" (Einstein 1931, 12). Spinoza used similar examples both in his *Ethics*, e.g., in part III, comment on theorem II, and in correspondence, e.g., in a 1674 letter to Schuller. (I am grateful to Hanoch Gutfreund and Elhanan Yakira for these references.) Arguments of this kind were also popular with the Stoics. Einstein often acknowledged Spinoza's influence on his worldview.

of reasons and final causes. The distinction between reasons and causes, and the repudiation of the former as an explanatory category in science, eventually ensued from one of the scientific revolution's major metaphysical achievements: the disentanglement of science and theology. Although many aspects of the early debates over the role of reasons in science now seem moot, others are still relevant to our own understanding of causal concepts and causal explanations in science.

The principle of sufficient reason well illustrates the intricate relations between reasons and causes in the seventeenth theory. Precursors of the principle can be found as far back as the earliest extant philosophical and scientific writings, but the principle's fame, and its name, are due to Leibniz, who called it "my principle,"[3] and took it to constitute the basis of every contingent truth.

> In order to proceed from mathematics to natural philosophy, another principle is required, as I have observed in my *Theodicy*; I mean the principle of sufficient reason, namely that nothing happens without a reason why it should be so rather than otherwise. (*Leibniz-Clarke Correspondence*, Leibniz [1915–16] 1956, second letter).

Leibniz uses a formulation of the principle in terms of "reasons"—*nihil est sine ratione*—interchangeably with a formulation in terms of "causes"—*nihil est sine causa*. By the latter, however, he is referring to *final* causes, that is, rational reasons as opposed to efficient or mechanical causes. Hence the formulation in terms of reasons and the formulation in terms of (final) causes, are one and the same. Yet Leibniz does occasionally have *efficient* causes in mind when referring to his principle. He maintained (as will be described in greater detail later in the chapter) that while explanations in terms of reasons are superior to explanations in terms of efficient causes, the two modes are complementary in that every natural phenomenon has a reason—a final cause—as well as an efficient or mechanical cause. Gradually, however, reasons and final causes, even where the term *reason* was used, were no longer taken literally. The principle of sufficient reason coalesced with the principle that

3. E.g., in *Leibniz-Clarke Correspondence*, Leibniz [1915–16] (1956), fifth letter.

every event has a cause—the causal principle.[4] Thus, when Laplace speaks of sufficient *reason*, he actually means sufficient *cause*.

> Present events are connected with preceding ones by a tie based upon the evident principle that a thing cannot occur without a cause [*cause*] which produces it. This axiom (is) known by the name of *the principle of sufficient reason* [*principe de la raison suffisante*]. (Laplace [1814] 1994, 3–4)

At the time, the questions raised regarding the principle of sufficient reason targeted both its validity and its meaning. The validity problem need not concern us at the moment,[5] but the principle's meaning merits consideration. The concept of a reason meant different things to different thinkers and underwent significant change over time. What, exactly, is a *reason*? Does it presuppose a reasoning mind? If not, if reasons are just certain types of natural connections among events or phenomena, can reasons still be distinguished from causes? Further, can an arbitrary will, human or divine, be considered a reason, or, for that matter, a cause? If we sanction such arbitrariness, are we thereby also committed to the possibility of chance? And what is a *sufficient* reason, how does it differ from a reason *simpliciter*? Aren't reasons analogous to causes, and thus, arguably, sufficient by their very nature? Understood as shaped by a reasoning mind, reasons, it seems, can indeed be good or bad, sufficient or insufficient, but construed as impersonal and natural, in what sense can such reasons be *in*sufficient? These and similar issues were of paramount importance to Descartes, Spinoza, Newton, Leibniz, and their contemporaries. We must, of course, remember that the metaphysical context of these seventeenth-century debates was very different from the classical context. For Aristotle and other classical thinkers, final causes do not presuppose a conscious mind that dictates the goal and progression of natural processes. Final causes and the attendant teleology are as much part of the natural order as are efficient causes.

4. As we saw in chapter 1, this is one version of the causal principle, the other being the "same cause, same effect" principle.

5. On the validity problem, see Pruss (2006).

In other words, on the classical understanding, while reasons, in the ordinary sense of the term, were considered to be final causes, final causes operating in nature were not necessarily conceived as reasons. In the Christian context of the scientific revolution, however, teleology was rooted in monotheistic theology, hence the reasons referred to in the principle of sufficient reason were God's reasons, or at least derived from them. In this context, the classical notion of a final cause—natural and mundane—no longer made sense. Yet final causes did not disappear; they survived because they became associated with rational reasons, that is, God's rational reasons.

Under the broad theological umbrella of the seventeenth century, different thinkers had radically different conceptions of God, and thus radically different positions on the possibility of ascribing reasons to God. The least personal of the various concepts of God was Spinoza's *Deus sive Natura*. Since such an impersonal deity could have neither reasons nor goals, final causes and teleological explanations were completely excluded from Spinoza's metaphysics and science. Necessity, whether causal or logical—for Spinoza, these categories are basically identical—is the only metaphysical glue.

Descartes's conception of God is more traditional: divine reasons exist, but are forever hidden from the human mind.

> And it would be the highest of presumption if we were to imagine that all things were created by God for our benefit alone, or even to suppose that the power of our minds can grasp the ends which he set before himself in creating the universe. ([1644] *Principles* III: 81; 1985, 1: 248)

Moreover, Descartes's God is free in the sense that even what we take to be necessary and unassailable truths of logic, mathematics, or metaphysics result from God's choice; that is, they are necessary only from the limited human perspective. The status of natural laws is similar: they too are decreed by God. This conception of the laws of nature had far-reaching implications for scientific method, as it rendered them humanly inexplicable. Being impenetrable, God's reasons cannot be invoked in scientific explanation. Matter, motion, and the natural laws that

govern them must suffice for a scientific account of natural phenomena, which must therefore be causal rather than teleological.[6]

> When dealing with natural things, we will, then, never derive any explanations from the purposes which God or nature may have had in view when creating them [and we shall entirely banish from our philosophy the search for final causes]. For we should not be so arrogant as to suppose that we can share in God's plans. ([1644] *Principles* I: 28; 1985, 1: 202)[7]

Despite their very different conceptions of God, then, Spinoza and Descartes reached the same methodological conclusion: both envisioned a science purged of teleology.

Newton, whose decidedly personal God was more accessible than the God of Descartes, took a different approach. On the one hand, Newton's system constitutes a perfect model of nonteleological explanation by means of mathematical laws and initial conditions. On the other, Newton allowed himself recourse to God's wisdom, will, and benevolence, but only when seeking to explain facts that he took to outstrip the explanatory resources of his system. Newton acknowledged the limits of his own science, and, arguably, the limits of science in general. He carefully distinguished between that which he considered himself to have demonstrated and that which he saw as hypothetical or speculative. This distinction is manifest both in the *Optics*, where bold ideas he could not prove (e.g., the bending of light by gravity!) appear as Queries at the end of the book, and in the *Principia*, where he candidly admits, in the General Scholium, that he has no explanation for the law of universal gravitation, and thus refrains from speculating

6. Descartes and his followers continue to speak of *efficient* causes, stressing their legitimacy and contrasting them to final causes, but the meaning of this Aristotelian term has also changed. Although Descartes does use "efficient cause" in the Aristotelian sense in a number of places (e.g., in [1644] *Principles* I, 28; 1985, 1: 202, where God is said to be "the efficient cause of all things"), in general, explanations in terms of efficient causes are gradually relinquished in favor of explanations in terms of natural laws.

7. The translation in *Philosophical Writings* (1985) is from the original Latin text published in 1644; the brackets in this edition indicate insertions from the French version published three years later.

about its cause—his famous "*hypothesis non fingo*" declaration. Furthermore, Newton distinguished between laws and initial conditions. He maintained that whereas the laws, together with the initial conditions, permit the derivation of every other state of a mechanical system, the initial conditions themselves are generally not explicable by science. The initial conditions of the solar system, for example, the fact that the planetary orbits lie (approximately) in the same plane, are seen by Newton as ensuing from a divine choice rather than determined by the laws of nature. Other problems Newton was unable to solve to his satisfaction within his system, such as that of the solar system's stability, were also relegated to extrascientific explanation in terms of God's benevolent intervention.

Leibniz initially shared the Cartesian commitment to efficient, rather than final, causes, but he later came to the conclusion that causes have only limited explanatory power, and reasons—the genuine manifestations of God's wisdom—are essential for adequate scientific explanation. Unlike Newton, who believed that God's free and unconstrained will should be considered a reason, Leibniz conceived of God as bound by the principles of reason. Whatever happens therefore happens in accordance with these principles, and in particular, the principle of sufficient reason. Moreover, whereas Newton appealed to God to explain lacunae for which he had no scientific explanation, Leibniz denied the existence of such lacunae. He took the principle of sufficient reason to constitute an integral part of science and to apply to every phenomenon. This, in my view, is the major difference between Newton and Leibniz. For Leibniz, no fact falls outside the jurisdiction of science; every fact satisfies the principle of sufficient reason and can therefore be explained by it. Consequently, Leibniz sought illustrations of God's wisdom and applications of the sufficient reason principle *within* science. These scientific applications differ significantly from more general invocations of the principle, such as apologetic responses to the problem of evil; in light of their empirical import, scientific applications could not be dismissed as mere metaphysical aberrations, or satirized as in Voltaire's *Candide*. Notably, in employing the principle of sufficient reason as a scientific tool, Leibniz legitimized teleological thinking in

science, a way of thinking that, as we just saw, several of his contemporaries had deemed passé. At the same time, in focusing on reasons and final causes within science, Leibniz drew attention to explanatory dimensions of science that other natural philosophers had not been attentive to. In particular, he realized the importance of symmetries and extremal principles, distinguishing them from lower-level explanations, which he considered inferior in scientific status.

Consider one of Leibniz's examples:

> Archimedes . . . was obliged to make use of a particular case of the great principle of sufficient reason. He takes it for granted that if there is a balance in which everything is alike on both sides, and if equal weights are hung on the two ends of that balance, the whole will be at rest. That is because no reason can be given why one side should weigh down rather than another. (*Leibniz-Clarke Correspondence*, Leibniz [1915–16] 1956, second letter)

Note that although Leibniz speaks of reasons, Archimedes's law is not overtly teleological or goal directed. We could easily substitute "cause" for "reason" in this quotation without damaging the argument. Symmetry considerations of this kind (which, as we saw in chapter 5, abound in modern science) are certainly different from ordinary causes—pulling, heating, and so on—but it is not clear why they should be construed as reasons. The question of what exactly Leibniz had in mind in so construing them is therefore intriguing; we will address it shortly.

Teleology is more conspicuous in the other example Leibniz cites: Fermat's principle. In 1662, Fermat demonstrated that light travels along paths that take the least time. To derive the principle, Fermat assumed that light travels along the path for which resistance from the surrounding media is minimal, and used a method he developed for finding geometric trajectories whose transit times are maxima or minima. Although this assumption is not in itself teleological, the resulting trajectory of least time demonstrates, in Fermat's view, that nature acts in the simplest and most economical way. Remarkably, using his principle, Fermat was able to derive the laws of optics known at the time,

most importantly Snell's sine law (discovered some forty years earlier), according to which when light is refracted, the ratio between the sines of the angles of incidence and refraction is a constant that depends only on the respective natures of the two media.[8]

Fermat's principle appears tantalizingly teleological. How can light pick out the right path other than by some sort of calculation that takes the entire path into account? The very notion of a *path*, involving specific end points, seems to presuppose goal directedness. Indeed, the principle was criticized straightaway on account of its ostensible ascription of foreknowledge to nature, in violation of scientific standards. Cartesians in particular, we have seen, were committed to efficient causes as the only valid explanatory resource. Faithful to this commitment, Descartes proffered a different, nonteleological derivation of Snell's law. Leibniz, however, considered Descartes's derivation of Snell's law by way of efficient causes "not nearly as good" as Fermat's (Ariew and Garber 1989, 55).[9] Not surprisingly, Leibniz recognized in Fermat's vision of the rational conduct of nature a vision close to his own. For him, Fermat's principle of least time was an instance of the more general principle of sufficient reason and a perfect illustration of the general principle's scientific fruitfulness.

It would be a mistake, however, to see Leibniz's appreciation of Fermat's principle as based solely on its teleological appearance. It had a more compelling rationale, a rationale that applied not only to Fermat's ostensibly teleological principle, but also to the previous example, Archimedes's law, which, as noted, does not have the same teleological character. Leibniz drew a distinction between explaining *individual*

8. A few years later, Huygens showed how Fermat's theorem and Snell's law could be derived from his wave theory of light. Huygens's method is closely related to subsequent discoveries leading up to quantum mechanics, in particular, the discovery of wave interference.

9. The passage in which the remark appears is from the letter to Johann Bernoulli written around 1698. Leibniz was not alone in his negative opinion of Descartes's derivation; other contemporaries, as well as later scholars, including Ernst Mach, expressed similar views. It is not entirely clear whether Descartes learned about Snell's law from correspondence or had already discovered it independently a few years earlier. See Sabra (1967) for a discussion of the priority question and a favorable reading of Descartes's argument. See also Mahoney (1973) for the details of Fermat's discovery and its reception.

events and explaining the *laws* that govern them. Individual events can be explained by ordinary causal laws, but the more ambitious goal of explaining those laws calls for the principle of sufficient reason. "I grant that particular effects of nature could and should be explained mechanically. . . . But the general principles of physics, and even of mechanics depend on the conduct of a sovereign intelligence, and cannot be explained without taking it into consideration" (Strickland 2006, 134).[10] It is not difficult to see why Leibniz thought that Fermat's principle provided an account of natural phenomena that was not merely law based, but rational and intellectually satisfying as well. The principle delineates the actual path of a light ray as distinguished from all other possible paths; it is unique by virtue of being *minimal*. This mathematical property is easily rendered in more metaphysical language as God's, or nature's, tendency toward maximal simplicity and economy. But even without the metaphysical gloss, the question "why this law, rather than some other law?" seems to receive a satisfactory answer, a "sufficient reason." Moreover, it is the sort of reason that can be dissociated from a reasoning mind and is thus closer to the classic *telos* than to conscious goal-directed action. On this reading, the difference between Leibniz and his opponents lies less in his endorsement of goal directedness than in his insistence on the unique character of higher-level laws. Recall that according to the deductive-nomological model of explanation, the most general laws, those that are not deducible from other laws, remain unexplained. Newton, as previously noted, acknowledged this limitation, conceding that the general law of gravitation may not have a scientific explanation. By contrast, Leibniz contended that the higher a law ranks within the scientific hierarchy, that is, the more general it is, the more elegant and rational it must be. For him, the most general laws are distinctively self-evident and self-explanatory; not only brute facts, but also inexplicable laws, have no place in the Leibnizian worldview.

10. This passage is from a note on the laws of nature (originally in French) apparently prepared in 1687 as a response to Malebranche. Strickland translated it from Leibniz's *Sämtliche Schriften und Briefe*, published by the Berlin-Brandenburgischen Akademie der Wissenschaften. For a similar expression of the need to explain both particular and general phenomena, see Ariew and Garber (1989), 283.

That science's explanatory ambition extends beyond the explanation of facts to the explanation of the laws is often noted by present-day physicists. As Richard Feynman put it, "in the further development of science we want more than just a formula. First we have an observation, then we have numbers that we measure, then we have a law which summarizes all the numbers. But the real *glory* of science is that *we can find a way of thinking* such that the law is *evident*" (1963, I, 26-3, emphasis in original). One way of rendering a law evident is by deriving it from a higher-level principle—a fundamental symmetry or a variation principle. Similar distinctions between lower- and higher-level laws were drawn, as we saw in chapter 5, by Wigner, who noted that symmetry principles are constraints on lower-level laws, and by Hermann Weyl, who maintained that symmetries, in contrast to ordinary laws, are given to us a priori. Like Leibniz, then, these thinkers invoke an explanatory hierarchy in which higher-level laws are epistemically superior to lower ones. Although they do not use the dated language of sufficient reason, the appeal of the desideratum that the most general laws should be the most self-evident has not waned.

Nearly a century separates Fermat's principle, the first "maxima minima" principle, from the formulation of the principle of least action. In the form given to it by Hamilton, and considered, as we will see, the principle's mature form, the least action principle asserts, roughly, that a mechanical system moving under the influence of conservative forces takes the path for which the action has an extremal point—a minimum or maximum (initially, only the first alternative was discovered, hence the principle's name).[11] More accurately, the system takes a path such

11. A force is said to be conservative when it is derivable from a scalar potential, or equivalently, when the work done around a closed orbit is zero. This means that there are no friction or other dissipative forces (whose work is positive) present. The conservative forces restriction is not a serious limitation with regard to fundamental forces, since these forces are all taken to be conservative. Nonconservative forces are introduced when we move from the fundamental level to higher levels. The least action principle is extendable to nonconservative forces; see Goldstein (1950, chap. 2). Although sometimes referred to as Hamilton's (first) principle, the least action principle should not be confused with Hamilton's equations. The least action principle represents the system's path from its position at time t_1 to another position at time t_2. It is thus a path in configuration space. Since only the initial position (and not the initial momen-

that for a small variation of the path, the first-order change in the action vanishes.[12] When the principle was first formulated, the definition of the action was the subject of bitter controversy, and is actually a thorny problem in general (more on this problem follows). In classical mechanics, the action is the integral over time of the difference between the kinetic and potential energies ($K - P$), or, since that difference is known as the Lagrangian—L—the integral of the Lagrangian over time. Thus, the action is

$$\int_{t_1}^{t_2} L dt = \int_{t_1}^{t_2} (K\text{-}P)\, dt$$

and the principle says that the system moves along a path such that, for a small variation of the path, the first-order change in this integral is zero. The principle of least action bears the same mark of apparent teleology as Fermat's principle of least time and raises the same question: how can a mechanical system "choose" the right path?

In terms of its teleological character, the principle of least action stands in stark contrast to Newton's laws. Recall that Newtonian mechanics was formulated in paradigmatically causal terms—forces acting on material particles and generating their acceleration.[13] The tremendous success of Newton's system and the intuitive appeal of its underlying causal picture played a decisive role in transforming the nature of scientific explanation from teleological to causal. The principle of least

tum) is given, determination of the path requires two different points. Hamilton's equations represent the path in phase space, where each point represents the system's position and momentum at the same time. Here, one point in phase space is sufficient to determine the path.

12. Note that higher orders (in the Taylor expansion) of the change in the action need not, in general, be zero. The variation refers to the difference between the actual path and alternative virtual paths between the same end points. The variation is expressed as a function of some parameter whose values correspond to each of the alternative paths.

13. The fact that, according to Newton, forces in general, and gravitation in particular, can act "at a distance," violating the desideratum that causal influence should be contiguous or (what we now call) local, was criticized by various thinkers, e.g., the *philosophe* d'Alembert. Newton's system was nevertheless widely viewed as a paradigm of causal explanation. As mentioned in chapter 1 herein, Einstein, whose commitment to locality is unquestionable, applauded Newton for creating the mathematical tools that best satisfy the physicist's demand for causality (Einstein [1927] [1954], 255).

action could not offer an equally intuitive causal model, and worse, it could easily be construed as reintroducing the teleological language that Newton's system had eschewed. Eventually, the appearance of teleology was dispelled when it was discovered that, under a wide range of conditions, Newton's differential formulation, which satisfies our causal intuitions, and the integral formulation of the least action principle, which ignores them, are in fact *equivalent*. For example, in the simple case of a particle moving with velocity v in a gravitational field whose potential is V, the force is dV/dx and the kinetic energy is ½ $mv^2 =$ ½ mdx/dt. The path that satisfies the least action principle is the very path that satisfies the differential equation

$$md^2x/dt^2 - dV/dx = 0$$

which is none other than Newton's second law!

The puzzle of how the particle picks the right path is resolved by noting that at every specific point along the path, it "gets directions" as to how to proceed, so that it does not need to "calculate" or "take into account" the entire path or its end point, let alone "choose" between alternatives. The teleological picture turned out to be no more than a superstructure resting on a more satisfactory causal structure.

The equivalence between these two formulations of mechanics was not apparent immediately. When Pierre-Louis Moreau de Maupertuis announced his least action principle in 1744, he was convinced he had discovered a *new* fundamental principle, not a reformulation of an old one. The merit of the new principle, in his view, was its ability to explain motion in terms of *final* causes. Ironically, this discovery was not an attempt to satisfy the Leibnizian principle of sufficient reason. On the contrary, it was formulated in the course of seeking to advance the Newtonian tradition in France. Maupertuis, who had established himself as a follower, popularizer, and translator of Newton into French, was critical of Leibniz's outlook in general, and of teleological interpretations of natural phenomena in particular.[14] He contrasted the proba-

14. Maupertuis was, however, very familiar with the Leibnizian tradition: he had studied

tive force of his new principle of least action with more familiar teleological arguments, such as the argument from design, which he deplored. It was a common mistake, he claimed, to point to some unexplained phenomenon, such as the structure of animals' limbs, or the initial conditions of the solar system, which had intrigued Newton, and then argue that nothing other than divine wisdom could possibly explain it. To expose the weakness of this kind of argument from design, Maupertuis devised causal explanations of the said phenomena. He put forward a schematic cosmological model that explained the solar system's puzzling initial conditions, and to account for animal limbs, an evolutionary conjecture that anticipated Darwin's natural selection. In his *Essaie de Cosmologie* (1750), he explained that a vast number of creatures were created by chance, but as only a few of them were well-enough structured and organized to survive, the majority perished. It was thus blind chance, rather than God's design, that produced the living world as we know it.[15] Maupertuis further argued that even granting that science cannot explain every phenomenon, the appeal to design in such exceptional situations does not amount to the kind of general explanatory theory found in science. By contrast, the principle of least action is a general mathematical law that leads to new predictions! It does not call upon God to fill the gaps in scientific explanation but, on the contrary, constitutes a scientific explanation so perfect that it bears the hallmark of divinity. According to Maupertuis, then, the principle of least action constitutes a triumph of legitimate teleological reasoning and at the same time exposes the flaws of traditional, unscientific teleological reasoning.[16]

---

with John Bernoulli, a disciple of Leibniz, and had a close relationship with Euler, who was also well versed in that tradition.

15. See, e.g., Harris (1981, 107) for a translation of the passages from Maupertuis (1750) that anticipate this Darwinian idea.

16. Although Maupertuis was trying to distance himself from Leibniz (e.g., he rejected the conservation of vis viva), from the perspective of his contemporaries, the Leibnizian flavor of the least action principle was unmistakable. So much so, that Maupertuis's priority was challenged by König, who credited Leibniz with the discovery. König claimed to have seen a letter written by Leibniz that stated the principle, but could not produce the original. The case was investigated by the Prussian Academy of Sciences, which decided in favor of Maupertuis (its

In fact, however, Maupertuis did not derive any new results from his principle. He applied the principle to simple examples—Fermat's principle, collisions of two particles and the law of the lever—in all of which he reached familiar, well-established results. Moreover, struggling with the definition of the action, he had to adjust this magnitude in each case to get the right result. But the modesty of these preliminary achievements should not detract from our appreciation of the conceptual innovativeness of Maupertuis's principle and its impact on the evolution of physics.

Euler, whose mathematical ingenuity exceeded Maupertuis's, gave the principle a more rigorous mathematical formulation and a wider range of applications, including many-body systems moving under the influence of external forces.[17] Yet unlike Maupertuis, Euler maintained that the principle only explains by different means effects that are also explicable by way of "efficient causes"—namely, Newtonian differential equations. Moreover, Euler considered the Newtonian route methodologically safer. The principle of least action, he said, can only be used "a posteriori" to recover the results derived from the differential laws of motion; it is not reliable as an "a priori" tool of prediction. The root of this uncertainty is again the concept of the action, which did not seem to have a strong intuitive basis and was often tailored to yield the desired outcome. Euler's assessment was vindicated by later developments—the action has indeed been redefined in every major transition to a new theory, a task that in the case of quantum mechanics, for example, was far from trivial. Euler was wrong, though, about the least action principle's utility as a tool of discovery.

Despite his reluctance to deem the principle of least action a new law of nature, Euler initially agreed with Maupertuis about its teleological

president at the time). König was accused of forging the letter, a conclusion that was disputed by the Academy in the following century. Today, the received view is that although König did not have, or even see, the original, the copy he cited was authentic. Even so, it appears that the letter did not contain a precise formulation of the least action principle and does not impugn Maupertuis's priority. Be that as it may, it was clearly the Leibnizian school of Bernoulli, Euler, and their followers that took interest in the principle and sought to further develop it.

17. Euler may have arrived at the principle a little earlier than Maupertuis did, but nonetheless attributed it to him, possibly because in their correspondence, Maupertuis had already stressed the importance of maxima minima principles.

significance. In his 1744 treatise on the calculus of variations, he expressed his conviction that nothing happens without a final cause that is reflected in some maximal or minimal property. Moreover, because natural effects can be derived from both efficient and final causes, they attest to the perfection of Creation and the wisdom of the Creator. Euler may have changed his mind on this reading of the principle, for a less teleological outlook, emphasizing the minimal magnitude of the cause of motion rather than the operation of a final cause, appears some twenty-five years later in one of Euler's *Letters to a German Princess*:

> You will find here, therefore, beyond all expectation, the foundation of the system of the late *Mr. de Maupertuis*, so much cried up by some, and so violently attacked by others. His principle is that of the least possible action; by which he means, that in all the changes which happen in nature, the cause which produces them is the least that can be. (Euler [1770] 1833, 265)

In the 1760s, Lagrange, inspired by Euler, began working on the calculus of variation and its application to mechanics. His formulation of Newtonian mechanics is the basis for much of what we now consider classical mechanics. Lagrange criticized Maupertuis's reasoning, in particular, his tenuous definition of the action. Although he did not adjust the principle's name, Lagrange recognized that it allows for a system's path to be maximal as well as minimal. Between Euler and Lagrange, the principle shed its metaphysical gloss. Euler may have wavered on the teleological interpretation, but Lagrange explicitly objected to it: "I view [the principle] not as a metaphysical principle but as a simple and general result of the laws of mechanics" (Lagrange [1811] 1997, 183).

With Hamilton, in the 1830s, the principle attained its mature form, cited above, in terms of the integral of the Lagrangian—namely, the path of a system moving under conservative forces is such that, for a small variation of the path, the first-order change in this integral vanishes. This variation principle is considered the apex of classical mechanics, and has retained its centrality in subsequent physical theories, such as the general theory of relativity and quantum mechanics. Since it uses only a system's potential and kinetic energies, both of which are scalar functions, the principle has the advantage of being independent

of the coordinates in which the Lagrangian is represented. This independence facilitates calculation, but more importantly, has gained theoretical significance due to Einstein's postulation of general covariance as a desideratum that physical theories must satisfy. Being invariant under coordinate transformations (as scalar functions are), the action, as well as functions and equations expressed in terms of the action, automatically satisfies this requirement. Moreover, the equivalence between the differential and integral formulations of mechanics has also received a firmer grounding, upon the demonstration that Hamilton's variation principle follows from the Euler-Lagrange differential equations and vice versa—the differential equations can be derived from the integral principle.[18]

Despite the mathematical equivalence between the differential and integral formulations of mechanics, the teleological reading of the latter remained alive even among prominent twentieth-century physicists. As late as 1937, for instance, Max Planck wrote:

> The least-action principle introduces a completely new idea into the concept of causality: The *causa efficiens*, which operates from the present into the future and makes future situations appear as determined by earlier ones, is joined by the *causa finalis* for which, inversely, the future—namely, a definite goal—serves as the premise from which there can be deduced the development of the processes which lead to this goal. ([1937] 1949, 179–80)

Planck was, of course, aware that "so long as we confine ourselves to the realm of physics, these alternative points of view are merely different mathematical expressions for one and the same fact" (180), but still marveled at the fact that "the most adequate formulation of this law creates the impression in every unbiased mind that nature is ruled by a rational, purposive will" (177). Planck's teleological reading of the least action principle was atypical for his time, but not as outlandish as it might at first seem. For if the two formulations of mechanics are indeed *equivalent*, could it not be argued that the causal level's being a mere

18. See, e.g., Goldstein (1950), chap. 2.

"superstructure" atop the teleological level was just as plausible as the reverse? If so, rejecting the teleological reading in favor of the causal one reflects a metaphysical or methodological preference, not the verdict of logic.[19] Debates over the best explanation often reach this sort of stalemate: both parties invoke their intuitions about explanatory force, and no further progress can be made, at least not within the old framework. Things start moving again when we get to the next round—quantum mechanics.

To the extent that the predictions of quantum mechanics are given in terms of *probabilities*, this poses a challenge to both teleology and determinism. Particles moving in accordance with quantum mechanical laws are not expected to follow a classical path at all, let alone the privileged path determined by the principle of least action.[20] And yet the predictions of quantum mechanics should converge on those of classical mechanics in the classical limit. How can we explain the emergence, from the underlying quantum world, which seems so different from the classical world, of the very special path that satisfies the least action principle? Richard Feynman gave an ingenious answer based on his path integral approach to quantum mechanics. In his work on classical electrodynamics, Feynman had already used the Lagrangian formalism rather than the Hamiltonian.[21] In moving to quantum mechanics,

19. Dirac, e.g., considered the Lagrangian formulation deeper than the Hamiltonian, though, as will be explained shortly, not for metaphysical reasons.

20. Even ignoring the indeterminacy of measurement results, the position—momentum uncertainty relations rule out the classical notion of a path, where each point represents a well-defined position and well-defined momentum.

21. In his doctoral dissertation, Feynman had identified two problems: the infinite self-energy of the accelerating electron and the infinite degrees of freedom of the surrounding field. Feynman sought to eliminate both kinds of infinity by assuming, first, that there is no self-interaction of the electron; and second, that there is no field. The theory he proposed was purely corpuscular—particles acting on each other. In the absence of fields, it was an "action at a distance" theory, but it satisfied relativistic locality in the sense that the mutual interaction of electrons was not instantaneous. That interaction could be expressed in terms of a field, but the field had no independent degrees of freedom other than those ensuing from particle to particle interaction. Wheeler (Feynman's dissertation advisor) and Feynman then showed that to get the correct results—i.e., no infinities—one had to assume that the absorber—the surrounding particles—responds by acting on the source by means of "advanced waves" (waves that appear

however, preference for the Lagrangian formalism was problematic, since up to that point, the equations of quantum mechanics had been formulated as a natural extension of the Hamiltonian equations of classical mechanics. The Lagrangian formalism did not seem amenable to an equally intuitive extension to the quantum domain, yet such an extension was required, Feynman believed, for creation of a quantum electrodynamics that paralleled his Lagrangian approach to classical electrodynamics. He therefore sought a quantum analogue of the classical action. The missing link was provided by Dirac, who in 1933 had published a paper titled "The Lagrangian in Quantum Mechanics." As it appeared in the outlying *Physikalische Zeitschrift der Sowjetunion*, it had received scant attention. Like Feynman, Dirac had come to the conclusion that it was imperative to find a way to formulate quantum mechanics in terms of the Lagrangian. Despite the close connection between the Hamiltonian and Lagrangian formulations of classical mechanics, there were, Dirac felt, "reasons for believing that the Lagrangian one is the more fundamental" ([1933] 2005, 111):

> In the first place the Lagrangian method allows one to collect together all the equations of motion and express them as the stationary property of a certain action function. (This action function is just the time-integral of the Lagrangian.) There is no corresponding action

---

to go backward in time), a solution to Maxwell's equation generally thought to be merely mathematical, with no physical realization. Wheeler and Feynman couched their theory in the Lagrangian formalism, invoking the principle of least action. As we already saw, this formulation, in contrast to Hamilton's equations, which describe the system's development over time, describes the entire path of a particle (system) between two end points. To Wheeler and Feynman, the advantage of this formulation seemed to be connected to the corpuscular picture they favored, for as Feynman put it later, the field variables were only "bookkeeping variables to keep track of what the particle[s] did in the past" (Feynman 1966, 36). For electrons, what makes the Lagrangian approach more suitable than the Hamiltonian is that the path of an electron at a given time depends on the paths of other electrons at *other* times. The Hamilton equations, which involve positions and momenta at the same time, are therefore unsuitable. Throughout his career, Feynman remained faithful to the Lagrangian formalism. In endeavoring to construct quantum electrodynamics along the lines that worked so well in the classical context, he was seeking to define a quantum action that was the closest analogue of the classical action. Precisely at this point, he happened to learn of Dirac's virtually unknown 1933 paper.

> principle in terms of the coordinates and momenta of the Hamiltonian theory. Secondly the Lagrangian method can easily be expressed relativistically, on account of the action function being a relativistic invariant; while the Hamiltonian method is essentially nonrelativistic in form, since it marks out a particular time variable as the canonical conjugate of the Hamiltonian function. ([1933] 2005, 111)

Dirac proposed a "correspondence" between a certain function of the classical action and the quantum mechanical transformation matrix, which governs the wave function's transition from one instant to another.[22] Feynman seized upon this clue, developing it into a novel picture of quantum mechanics that did not start out from the existing formalisms (i.e., those of Heisenberg, Schrödinger, and Dirac), and proved extremely useful in creating quantum electrodynamics. Using the Lagrangian proposed by Dirac, Feynman recast quantum mechanics in terms of path integrals, and derived the Schrödinger equation therefrom. The principle of least action turned out to be the classical limit of this novel path-integral rendering of quantum mechanics!

On Feynman's approach, quantum mechanics is an algorithm for calculating probability amplitudes. Incorporating Dirac's proposal into this picture entails that probability amplitudes have a *phase* proportional to the classical action and responsible for the periodicity characteristic of quantum phenomena such as interference. The nonclassical character of the theory manifests itself in the fact that, as we saw in chapter 4, probabilities interfere with each other in ways that deviate from classical expectations. Distinguishing interfering from noninterfering states is crucial for counting alternatives and calculating probabilities.[23] Feynman's idea was that the probability amplitude of an event is the sum of the amplitudes of every possible way in which the event can occur. Each possibility is represented by a path, and each path is assigned a probability amplitude. Thus, a photon traveling from a source to a par-

22. Dirac observed that *exp* $2\pi i/h$ multiplied by the classical action corresponds to the transformation matrix.

23. Since, on Feynman's approach, there is no underlying equation, formalism, or intuitive model that determines quantum probabilities, such counting is far from trivial, but for our purposes here, we can set this issue aside.

ticular point on a screen can reach that point by infinitely many paths, each one correlated with a probability amplitude. The total probability amplitude of the event—in this case, the photon's reaching the designated point—is the sum of the probability amplitudes assigned to the possible paths—namely, the integral of the amplitude over all possible paths.

In their attempt to "demystify quantum theory," Cox and Forshaw (2011, 4) endorse Feynman's version of quantum mechanics. Their title, *The Quantum Universe: Everything That Can Happen Does Happen*, epitomizes the path integral approach on which we integrate over all possibilities. At first sight, nothing seems farther from a causal conception of the world than the liberty to realize every possibility. On reflection, however, the disparity between the Cox-Forshaw aphorism and causation (broadly construed, as per this book) disappears. To see why, we need only consider the aphorism's contrapositive—everything that does not happen cannot happen. It mandates that we explain what does not happen by pointing to laws or constraints that exclude it. Pauli's principle, discussed in chapter 5, can serve as an example. We never find multiple electrons occupying the same quantum level. There must therefore be some constraint precluding this situation, as there indeed is: Pauli's exclusion principle. The merit of Feynman's approach is that its freedom-inducing orientation facilitates the derivation of such constraints. Not only does it recover Pauli's principle, it also explains and predicts with great precision many of the constraints on elementary particle interactions.

Rather than describing the state of the system at each moment as a function of its state at the previous moment, the path integral method, like the classical principle of least action, involves looking at paths in their entirety. Feynman maintained that the path integral approach was consonant both with the time-reversal symmetry of the fundamental laws of nature (since it could be seen as countenancing particles moving backward in time) and with Einstein-Minkowski four-dimensional spacetime (since it could be made Lorentz-invariant).[24] Besides, the

24. See Feynman (1963, I, 19); Feynman (1965), and the 1948 manuscript quoted in Schweber (1994, 432–33). One consequence of the path integral approach was that the continuous wave model could not be sustained.

path integral approach was beautiful: "The behavior of nature is determined by saying her whole spacetime path has a certain character" (Feynman 1965, 35), and Feynman, as he puts it, "fell deeply in love" with this approach (1965, 32).

How, then, does the method of integration over paths explain the emergence of the classical limit—the classical path obeying the principle of least action? As we saw, Feynman took the probability amplitude, represented by a complex number, to have a phase proportional to the classical action. The different paths that are integrated over thus have different phases. At the classical limit, the classical action is so large in proportion to *h* that in general, even a small variation of the action amounts to a large difference in the phase, expressed in quantum units. Paths (and probability amplitudes) that are close to one another in classical terms may thus still differ considerably in phase. These differences are reflected in the phase's rapidly changing its sign between nearby paths. Consequently, the majority of paths will be canceled out by paths with similar classical action but opposite phase. When we approach the stationary point, however, where a small variation of the path does not alter the (first order of the) action, the phase no longer oscillates, and amplitudes—paths—are not canceled out. On the contrary, here the different amplitudes add up, so that the probability amplitude is largest around the stationary point. We therefore see a large number of particles moving along this particular path—the classical path. It is here, for the first time, that the principle of least action is given an explanation that completely defeats the teleological interpretation that had accompanied it for more than two centuries. Furthermore, the emergence of the privileged classical path from the underlying multitude of quantum possibilities is a beautiful example of the general phenomenon of *emergence*, a concept fiercely debated in contemporary philosophy of science (and the subject of chapter 7).[25]

Our understanding of the principle of least action has thus taken another turn. In classical mechanics, the response to the charge of tele-

25. A common critique of Feynman's path integral approach is that it constitutes an ingenious mathematical technique, not a physical explanation. Even so, I would argue, the very possibility of accounting for the apparent teleology of the classical path by means of a formalism that altogether shuns teleology is eye-opening.

ology was, as we saw, that the system does not "reason" or "calculate," but simply follows the "instructions" expressed by the (Newtonian, Euler-Lagrange, or Hamiltonian) differential equations. The future-oriented tendency of a system moving in accordance with the principle turned out to be an illusion, to be merely apparent teleology. The quantum response to the putative teleology is more radical: the system does not even get instructions—it goes whichever way it goes, and no path is excluded in advance. What we, macroscopic creatures that we are, perceive as very special behavior is the result of a selection process by which many paths cancel each other out while others add up and reinforce one another. Selection processes of this kind tend to give us the impression that nature "prefers" certain possibilities to others and "chooses" accordingly, enticing us into teleological thinking.

We have already encountered apparent teleology arising from selection processes in earlier chapters of this book. We saw that when a specific result is stable or overdetermined, that is, insensitive to initial conditions and small perturbations, we tend to interpret it as the preferred outcome of the process in question. In such cases there is no actual selection process going on, but there is a kind of virtual selection underlying the realization that, had the initial conditions been different, we would still be getting the same stable result. In the theory of evolution, even this picture of virtual selection is inaccurate. There is, of course, much "canceling out," leading us to say, as we do, that the organisms that have survived have been selected. Yet the stability of the actual structure or behavior that has survived is not, in fact, guaranteed. In some cases, we can prove by game-theoretic techniques that certain structural or behavior patterns are indeed evolutionarily stable, but there is no general argument to prove that this is universally true with regard to every biological feature. Sometimes stability, like teleology, is only a post-factum projection on our part. By contrast, on Feynman's approach, the majority of quantum paths actually cancel each other out, and a "selected" trajectory emerges.

Feynman tells us that he was intrigued by the principle of least action from the time he learned about it from his high school physics teacher. The interpretation he came up with, apart from its renowned merits as

a physical theory, also provides an ingenious solution to the long-standing philosophical problem of teleology in physics. It is, perhaps, the greatest contribution to philosophy ever made by someone who had as much contempt for philosophy as Feynman. Resisting the temptation of teleology is an important aspect of what, in the wake of Max Weber and Friedrich Schiller, can be called the disenchantment of nature precipitated by modern science. What these thinkers could not have foreseen was that chance would be even more effective than the deterministic science of their day in bringing about this disenchantment.

7

Causation and Reduction

IN THE FOREGOING CHAPTERS, we have become acquainted with various causal constraints operative in physics, and their interrelations. In contrast to "republicans" (causal eliminativists) such as Russell and Norton, who seek to banish the concept of cause from fundamental physics, I have argued that causal constraints—constraints on possible change—have been the cornerstone of physics from classical mechanics to contemporary theories. It is now time to turn to another variant of causal eliminativism, a variant, mentioned only briefly in chapter 1, that can be described as the polar opposite of "republicanism." Proponents of this second sort of eliminativism acknowledge the centrality of causation at the fundamental level of physics, but maintain that this fundamental level is the *only* level where causal notions are applicable. Putative causal claims involving higher-level events are, according to these eliminativists, either reducible to the causal claims of fundamental physics, in which case they are redundant, or simply false. Let us call this position "higher-level eliminativism," to be distinguished from the "republican" denial of causation at the basic level. As it confines causation to the fundamental level, this position is sometimes referred to as "causal fundamentalism." Assessing the cogency of higher-level eliminativism involves scrutiny of the relations between different conceptual levels (or different levels of reality), that is, it involves clarification of the concept of reduction. The scope and limits of reduction are the subject of this chapter. We will see that not only is there a place for higher-level causation, but the possibility of lawlessness suggests that

there is also room for more radical deviation from the reductionist vision—namely, conceptual categories that completely resist subjection to projectable scientific laws.

The debate over higher-level causation is particularly pertinent to the philosophy of mind, where the causal efficacy of mental events and mental properties has long been questioned. But the debate is also relevant, as we will see, to events, properties, and concepts that are physical but are not present at the fundamental level. With regard to higher-level events and properties, mental as well as physical, the question at issue is whether they can be covered by the laws of fundamental physics. In a nutshell, the concern that motivates higher-level eliminativism is that causal relations at (or between) higher levels (or between higher levels and the fundamental level) threaten to disrupt the physical order at the fundamental level. Were such higher-level causal relations to exist, it is claimed, they would interfere with the causal autonomy—the physical closure—of the fundamental level. I will argue that this concern is unfounded.

To get started, it will be useful to review some familiar developments in twentieth-century philosophy of mind.[1] In *The Concept of Mind* (Ryle 1949), Gilbert Ryle argued that the mental is inseparable from the physical. The traditional view—"the double-life theory," as he put it (19)—which seeks to differentiate the mental realm from the physical, is based on a "philosophical myth," or worse, a logical error—"a category mistake" (17). The mind is not to be characterized by what it is made of, but rather by how it is organized. On Ryle's view, one can concede the difference between mental and physical activity without being committed to the existence of a mysterious mind stuff—"the ghost in the machine" (17), and without having to address the question of how such disparate kinds of entities, matter and mind stuff, could interact. To drive the point home, Ryle adduces the concept of a university as an analogy. Obviously, universities are housed in buildings, but it is learning and research, not buildings, that makes them universities. It would

1. This summary does not purport to be exhaustive but focuses on arguments that have bearing on the issues discussed here.

nonetheless be a category mistake to picture a university as a separate entity over and above its buildings. The mind theorist who takes the mental to be something that exists *in addition* to the physical is analogous, Ryle asserts, to a visitor who has seen all the buildings on campus but complains that she has not seen the university. Traditional conflicts between monists and dualists over the makeup of the mental are rendered obsolete by Ryle's approach.

Although Ryle did not focus on the problem of reduction, he addressed it in passing apropos his discussion of freedom of the will.

> The fear that theoretically minded persons have felt lest everything should turn out to be explicable by mechanical laws is a baseless fear. . . . Physicists may one day have found the answer to all physical questions, but not all questions are physical questions. (Ryle 1949, 74)

Unlike Ryle's mental/physical identity thesis, which was immediately recognized as a significant contribution, his penetrating observation about reduction had little impact. Among the next generation of philosophers, however, reduction became a central topic in both the philosophy of science and the philosophy of mind. Hilary Putnam's "Philosophy and Our Mental Life" pioneered the position that came to be known as functionalism (or machine functionalism). In the paper, Putnam, like Ryle, downplayed the issue of what the mind is made of, stressing its function and organization: "The question of the autonomy of our mental life . . . has nothing to do with that all too popular . . . question about matter or soul-stuff. We could be made of Swiss cheese and it wouldn't matter" ([1973] 1975b, 291). Hence in terms of ontology, understanding mental activity does not require a richer ontology than that mandated by our best physical theory. But this physicalist ontology, Putnam emphasized ([1967] 1975a), does not guarantee reduction. The reason Putnam gave for the failure of reduction, at this functionalist phase in his thinking, was that the same function could be realized by different physical systems, a feature known as multiple realizability.[2] As

2. The term was coined by Fodor in his 1974 "Special Sciences."

an analogy, Putnam invoked computers, where the same software is compatible with many kinds of hardware. Mental functions, or the computations corresponding to them, must be compatible with the laws and mechanisms of the different systems that realize them, but are not reducible to the workings of any one particular physical system and its laws. We therefore cannot identify a particular kind of functional mental state with a particular kind of physical state in the way that we identify light, say, with an electromagnetic wave. In the 1990s, in critiquing his earlier functionalism as overly reductionist, Putnam found further fault with reductionism. In line with his externalist theory of meaning, he now claimed that mental states—for instance, remembering last night's movie—often presuppose interaction with an external environment. Consequently, *inner* mental states, whether physical or functional, cannot provide an exhaustive account of meanings, intentions, or of the mental more generally. Mental states also have contextual, social, and individual aspects, such as figurative rather than literal meanings of terms, or the associations formed in individuals' minds through personal experiences. Externalism and contextuality render the reducibility of mental states to physical states unfeasible (Putnam 1994).[3]

In "Mental Events," Donald Davidson ([1970] 1980) launched a different, though related, attack on the reducibility of the mental to the physical. Davidson characterized mental events loosely, as events that have mental descriptions. They are thus physical events that can be described in mental terms, for instance, being delighted or surprised. Although he endorsed the identity thesis for individual mental events (token identity), taking every mental event to be a physical event, Davidson denied type identity: mental events of a certain type, say, being surprised, do not constitute a corresponding physical type. Underlying this denial is the observation he made in his seminal "Causal Relations" ([1967] 1980) regarding the description sensitivity of nomological explanations. Laws involve types fundamentally, as they

3. Putnam did not retract the core thesis of functionalism, that is, he in no way rejected the priority of function over material makeup. His later view differs from the earlier one in recognizing that there are stronger arguments against the mental's reducibility to the language of physics than multiple realizability.

invariably connect types of events rather than individual events. Types, in turn, are referred to via descriptions. In order for individual events to be subsumed under and explained by laws, they must be described in terms of the predicates appearing in those laws (the types connected by the laws are the extensions of these predicates). If we do not describe an event appropriately, that is, in the language of the laws, but rather use some alternative description, it may be impossible to predict or explain the event in question; the derivation of the event under that description may be blocked.[4]

Here again we see multiple realizability—a mental type is realizable by physical events that fall under multiple physical types. Davidson refers to this relationship between the two kinds of events as the supervenience of the mental on the physical. Worlds that differ in the mental states of their inhabitants necessarily differ in their physical states, but the converse need not hold, for worlds differing in their physical states may still exemplify the same mental states. The relation between physical and mental events is a many-one relation.

Combining these insights, Davidson managed to reconcile three assumptions that at first sight seem hopelessly at odds: the causal interaction of mental and physical events, the Humean understanding of causation in terms of law-like regularities, and the repudiation of laws couched in terms of mental predicates. The significance of description underpins this reconciliation. An individual event that is predictable and explicable by the laws of physics under one of its descriptions may elude prediction and explanation under numerous other descriptions, and in particular, under mental descriptions. Since mental types do not correspond to physical types, there may be no physical law (or set of physical laws) that invokes the mental types in question, either directly, or indirectly via their correspondence with physical types that are subject to law. And although an individual mental event can be the cause or the effect of a physical event, and although there is a law—a nomological connection—underlying any such causal relation, the law will refer to the mental event only under its physical description, not its

4. E.g., when we describe a free-falling object in terms of its initial height above the ground and its initial velocity, we can predict the velocity with which it hits the ground, whereas if we describe it in terms of its color and chemical structure, we cannot.

mental description. Mental events can therefore be covered by physical laws, though not by laws formulated in terms of mental types (or by mixed laws that connect mental types with physical types). The compatibility of the above assumptions is therefore saved. Davidson calls his position anomalous monism—it is monistic in the sense that the mental is physical, and anomalous in the sense that the mental does not fit into the web of physical laws. Another common name for this position is nonreductive physicalism.

I will uphold Davidson's approach, defending it against various objections, and articulating its implications for reductionism in areas other than the philosophy of mind. A number of caveats should, however, be mentioned. Davidson was committed to a Humean understanding of causation in terms of regularities, but his point regarding description sensitivity does not actually hinge on that commitment; it is sufficient to argue that *when* there are lawful regularities, they refer to types via particular descriptions. Granted, the Humean commitment renders Davidson's success at reconciling his seemingly conflicting assumptions all the more surprising, but it is not crucial for his main point, or for the use made of it here. Similarly, Davidson construes causal relations as relations between *events*, a restriction that is peripheral to what I take to be his key insight. In discussing examples from physics, it is convenient to speak of states, properties, and processes as standing in causal relations and subject to causal constraints. Lastly, whereas Davidson takes mental events, and only mental events, to be anomalous, my view is that the division between the lawful and the lawless does not coincide with the division between the physical and the mental. On the one hand, some mental events may well satisfy certain laws even under their mental description. On the other hand, there are concepts (and types of events) that, though not mental, cannot be captured by physical laws. Clearly, whether mental or not, concepts that lie outside the jurisdiction of physics pose a threat to the reductionist program.

Basically, there are two approaches to reduction, one in terms of the logical relations between theories, the other in terms of causation. The former, proposed by Ernst Nagel (1961, chap. 11), requires that the concepts of the reduced, higher-level theory be defined by means of fundamental-level concepts, and that the higher-level laws be derived

from the laws of the fundamental level.[5] In view of the paucity of examples that satisfy these strong requirements, they are often weakened in the following way. The definitions in question need not establish the synonymy of the defined (reduced) terms with the defining terms, but rather, the definitions can be empirical laws (bridge laws) establishing coextensionality rather than identity. And the laws derived from fundamental-level laws need not be identical to the laws of the reduced (higher-level) theory, but rather, it suffices that the latter constitute good-enough approximations of the fundamental laws. The fundamental laws can, for example, yield a probabilistic version of the laws of the reduced theory, as in the derivation of thermodynamics from statistical mechanics.[6]

The second approach to reduction, focusing on causation, takes reduction to show that underlying the causal relations at higher levels are causal relations at the fundamental level. When reduction of this kind is achieved, genuine causation exists only at the fundamental level. Given the lack of consensus on the meaning of causation, the causal approach to reduction is more ambiguous than the Nagelian approach. For instance, depending on whether we understand causation in terms of lawful regularities, the two approaches to reduction can be seen as competing or complementary. In any event, on both approaches, successful reduction makes higher-level theories (in principle, even if not in practice) redundant. Reduction*ism*, in turn (on either account), seeks to reduce *all* higher-level theories (and phenomena) to the most basic level of fundamental physics. Although the foregoing summary of reduction is highly schematic, it suffices to enable us to address the concerns that motivate higher-level eliminativism.[7]

5. This formulation may not be faithful to the letter of Nagel's account, but is consonant with its sprit. Note that I am only discussing what Nagel (1961, 342) refers to as "heterogeneous reduction."

6. Here I ignore the current debate about whether the reduction of thermodynamics to statistical mechanics has actually been achieved; see chapter 3 and the literature cited there.

7. As Nickles (1973) observed, on a different (and indeed, opposite) usage of the notion of reduction, common among physicists, it is the fundamental theory that is reduced to the higher-level theory, meaning that the former converges on the latter in the limit. Thus, whereas the philosopher would take Newtonian mechanics to be reducible to the special theory of relativity

To understand these concerns, let us distinguish between types of relations that may obtain between different levels, or between the laws operative at those levels. Consider a fundamental level *F* and a higher-level *H*. At the outset, we should note that the laws of *F* and the laws of *H* can be consistent or inconsistent with each other. As already mentioned, there are actually very few cases where higher-level theories are rigorously consistent with lower-level ones; typically, the laws of the basic level contradict those of the higher level.[8] But let us agree to settle for a weaker condition than perfect consistency—one theory's being consistent with a good-enough approximation of the other—and assume that this condition is satisfied in the case of *F* and *H*. There are still at least three possibilities:

1. Reduction: All *H*-laws can be reduced to *F*-laws, so that *H*-laws are eliminated in favor of *F*-laws. In this case *H*-laws are redundant and *H*-level phenomena are deemed epiphenomena.
2. Lacunae: There are *H*-laws that cover (predict and explain) phenomena that *F*-laws do not cover (lacunae).
3. Overdetermination: There are *H*-laws that are irreducible to *F*-laws but provide alternative predictions and explanations of phenomena that *F*-laws suffice to explain. Being entailed by two distinct sets of laws, these phenomena are thus overdetermined.

And similar relations can be formulated in terms of causality:

1. Reduction: All *H*-causes are actually *F*-causes, hence *H*-causes are redundant.

---

at velocities much lower than that of light ($v \ll c$), the physicist might say that special relativity reduces to Newtonian mechanics. I will use "reduction" in the philosophers' sense, which is more apt for discussing the problems that concern us here.

8. This is clearly the situation in statistical mechanics—the reductionists' favorite paradigm case—but it is also what happens in simpler cases that are usually thought of in terms of generalization rather than reduction. Strictly speaking, Newtonian mechanics contradicts Galileo's law of free fall, but the affinity between the two theories' respective predictions for small enough terrestrial distances induces us to think of Galileo's law as an instance of Newton's more general law.

2. Lacunae: Some *H*-causes bring about effects that have no *F*-cause.
3. Overdetermination: Some *H*-causes, though irreducible to *F*-causes (that is, though not identical to any *F*-cause), bring about effects that *F*-causes also suffice to bring about.

Denying the possibility of lacunae and overdetermination, reductionists see only the first option as viable. Their reasoning involves the deterministic assumption of the physical closure of the basic level: the assumption that every basic-level event is determined (predictable, explicable) by the laws and initial conditions (or boundary conditions) of that level.[9] This deterministic assumption only holds for closed systems and is valid only for classical theories, not quantum mechanics. Nevertheless, if, for argument's sake, the assumption of physical closure is accepted, lacunae and overdetermination are ruled out. Reductionism is vindicated, or so it seems.

Debates over reductionism are often intertwined with questions about emergence, a concept that is used in several different ways. Emergence is commonly characterized as simply the opposite of reduction. On this understanding, when there is an in-principle (rather than a merely practical) failure of reducibility, the irreducible phenomenon is emergent. There are also views, such as that expressed in Butterfield (2011), which take the emergent to be fully reducible, but maintain that, in comparison with basic-level phenomena, emergent phenomena and laws nonetheless manifest conceptual and phenomenological novelty.[10] And there are "mixed" views that characterize the emergent as reducible in one sense, say, in supervening on the fundamental level, but

9. As noted in chapter 2, determinism does not actually guarantee predictability, a point that will be addressed below.

10. According to Butterfield, emergent phenomena are reducible to the basic level in the sense of being derivable from its basic-level laws, albeit by special mathematical methods such as the renormalization group. Butterfield further argues that in such cases, the higher-level concepts are implicitly defined by lower-level ones, and that Beth's theorem, according to which implicit definitions can be made explicit, implies the definability of higher-level phenomena by means of lower-level ones. It is possible to accept Butterfield's understanding of emergence without committing ourselves to the strong claim about implicit definition.

not in another, for instance, not in the Nagelian sense of reduction. A particularly instructive such combination view is explored by Mark Bedau (2008) in support of what he calls "weak emergence." Drawing on complexity theory and its application to cellular automata (Wolfram 1994), and on the definition of randomness in Chaitin (1966), Bedau distinguishes between cases where there is a general law—a mathematical formula—directly connecting every state of a system with every other state, and cases where there is no such law. The law, he argues, supplies a "short cut" (Bedau 2008, 162) between distant states, so that, say, given the initial state, there is no need to go through every intermediate state in order to calculate the final state. Laws of this kind are the core of physical theory. Yet it can be shown that there are systems—and deterministic systems at that—whose evolution cannot be captured by such a law. When this is the case, the only way to establish which state the system should be in at a certain point is to let the system run its course until it reaches that point, or to run a computer simulation of the process, which is, in principle, equivalent (and obviously, more practical). Such systems are deterministic in the sense that runs, or simulations, with exactly the same initial conditions yield the very same trajectory, so that there is no randomness in the traditional sense of allowing an open future. Nevertheless, systems of this sort are random in an alternative sense of the term defined by Chaitin: they are lawless and unpredictable (except by simulation).[11] Examples of such systems are supplied by cellular automata, studied in great detail by Wolfram (1994) and described in Bedau (2008). In the Game of Life (Berlekamp, Conway, and Guy 1982), for example, every step is uniquely determined by the game's update function together with the configuration reached in the previous step;[12] the game is therefore deterministic. At the same

11. Chaitin views his discovery of randomness in this newly defined sense in the context of mathematical logic's other fundamental limit theorems: "In a nutshell, Gödel discovered incompleteness, Turing discovered uncomputability, and I discovered randomness—that's the amazing fact that some mathematical statements are true for no reason, they're true by accident. There can be no 'theory of everything,' at least not in mathematics" (Chaitin 1999, v).

12. The game is run on a two-dimensional configuration of cells that have two possible states, "dead" and "alive." An example of an update is as follows. A living cell at step n remains alive at step $n + 1$ if and only if, at step n, two or three of its neighbors were alive. A dead cell at step n

time, since for many initial configurations there are no short-cut laws, the game exhibits Bedau's "weak emergence." Understood in this sense, emergence presupposes no theoretical stratification and no division into higher and lower levels of reality.

Although there is no need to decide between the various senses of emergence, it is important to be aware of the differences between them. In what follows, I focus on reduction, first within physics, and then in general. The question of whether, and in what sense, emergence is tenable, will take care of itself once the scope and limits of reduction are clarified.[13]

Are there any examples of the failure of reduction in physics? We saw that in the context of the philosophy of mind, multiple realizability has been taken by Putnam and Davidson to suggest such failure. Note that neither of them actually demonstrated the multiple realizability of the mental; its role in their arguments is that of an assumption, or a conclusion derived from some other philosophical thesis, such as functionalism. Multiple realizability is, however, well established and quite common in physics.[14] In examining the case of statistical mechanics in chapter 3, we saw that macrostates are multiply realized by microstates. In particular, the entropy of a macrostate corresponds neither to a specific microproperty nor to an average over microproperties, but rather to the *number* of microstates (or the volume in phase space) belonging to that macrostate, and thus to the probability of the macrostate. The concept of entropy therefore involves the higher-level concept of macrostate essentially, that is, it cannot be defined solely in terms of microproperties. Furthermore, by singling out the *size* of a macrostate as its most significant physical property, this understanding of entropy

---

is revived at step $n + 1$ if and only if, at step n, it had three living neighbors. See Bedau (2008) for more detailed examples of the game under varying update functions and initial conditions, and their computer simulations.

13. If, for example, there are no failures of reduction within physics, there will be no emergence (in the first sense) in physics; if there are such failures, Butterfield's concept of emergence will require modification.

14. Scientists do not often use this philosophical term, but are aware of the many-one relation to which it refers.

highlights the remarkable indifference of macro phenomena to the physical properties of individual microstates. Consequently, statistical-mechanical explanations of macro phenomena such as the stability of one macrostate—equilibrium—relative to others, or the limit on the efficiency of heat engines, are not based solely on laws operative at the fundamental level, but require higher-level concepts and laws.

The multiple realizability of macrostates in statistical mechanics illustrates Davidson's observation that the various descriptions of an event determine the types it can belong to and therefore also determine the laws under which it can be subsumed. It makes perfect sense for an event to instantiate the laws of physics under one of its descriptions and fail to do so under others. And we may now add that it also makes sense for an event to instantiate one law under one of its descriptions and a different law under an alternative description (provided these laws are consistent with each other), an option Davidson did not consider.[15] Consider a system that, at a certain moment, is in a microstate belonging to the equilibrium macrostate. If characterized as a type of microstate, that is (*per impossibile*), in terms of the precise positions and momenta of its trillions of particles, that microstate and the subsequent evolution of the system can be subsumed under the laws of mechanics. But on its own, this description will not tell us anything about, let alone explain, the system's macrostate; for instance, it won't tell us that the system is in a state of equilibrium. To explain the stability of this particular macrostate and the ramifications of this stability for the system's subsequent development, we must adduce the system under its macroproperty description, and the laws operative at the macrolevel. In other words, the behavior of macrostates qua macrostates cannot be explained by the fundamental laws of the microlevel. By the same token, causal constraints on macrolevel processes, constraints that determine which macrolevel changes are more probable than others, are additional to the causal constraints characteristic of the microlevel. That there are

15. The consistency of statistical mechanics with the fundamental laws of physics remains an open problem, but here we can assume that it is solvable (see the end of chapter 3). The problem pertains primarily to the directionality of the second law of thermodynamics, an issue that is not crucial for the present argument.

such additional constraints does not attest to overdetermination of microevents or to any lacunae at the microlevel. Rather, the additional laws and constraints involve new, higher-level, types, about which the fundamental laws are silent. As long as the laws applicable to these new types are consistent with the fundamental laws, there is no violation of the physical closure of the fundamental level.[16] These considerations, I should stress, remain valid even if macrostates supervene on microstates—namely, even if no change in the system's macrostate can occur without change in its microstate as well. (It was noted in chapter 3 that supervenience holds in Boltzmann's statistical mechanics, but not Gibbs's.) Supervenience ensures that any transition from one macrostate to another is, ipso facto, also a transition from one microstate to another. It also ensures that every microstate belongs to a single macrostate, so that if we could identify the two microstates in question, the identity of the corresponding macrostates would also be fixed. Yet without the additional information about the relative size of these macrostates—information that is foreign to the microlevel—no explanation of their behavior qua macrostates can be gleaned from the laws of the fundamental level.

It is often thought that the fact that each macrostate is realized by a microstate suffices to establish the reducibility of macrostates. But this is a category mistake like those Ryle cautioned us about. A system that is in a particular macrostate (at a particular moment) is also in a particular microstate, that is, it instantiates both a macrostate and a microstate, but this identity falls short of reduction in both the logical and the causal senses of the term. (Although every mother is a woman, motherhood is not reducible to womanhood.) Insofar as reduction pertains to laws operative on macrostates, no reduction is achieved by pinpointing the microstates that realize them. Insofar as it pertains to causal relations at the macrolevel, the instantiating microstate in itself is likewise inert.

16. The example about shuffling a deck of cards, discussed in chapter 3, illustrates the same point. Individual series are equi-probable, but under the higher-level concepts of ordered and disordered decks, we can explain why disordered decks are more probable. This explanation does not invalidate or render superfluous the detailed explanation of how, by means of a number of specific steps, we get from one particular series to another.

If, for example, the stability of a macrostate is considered causally relevant to the macrostate's response to small perturbations, this causal efficacy cannot be ascribed to the particular microstate that happens to instantiate the stable macrostate at a particular moment. The fact that a small perturbation would alter the microstate, while most probably leaving the system in the same macrostate, is essential to our understanding of macro phenomena. Such ascription of causal efficacy to macrostates does not entail that there are lacunae at the microlevel, or deficits in its physical closure. Concern about overdetermination is likewise misplaced. Macrostates are indeed insensitive to the precise nature of their realizing microstates; numerous *other* microstates would have produced the very same macrobehavior. This overdetermination, however, is not present at the level of microstates and microevents; only macrostates are multiply realizable, and thus overdetermined, in this way. The apprehensiveness regarding an alleged incompatibility between macrolevel causality and the physical closure of the fundamental level is, again, unwarranted.

Another example of multiple realizability in physics is the phenomenon known as universality: the strikingly similar behavior of very different physical systems at (or close to) specific points—*critical* points. Universality is frequently cited as attesting to multiple realizability, as well as emergence, and has been analyzed in detail (e.g., Batterman 2002; Butterfield 2011). Water and ferromagnetic materials have little in common in terms of physical/chemical structure and behavior. But during phase transitions such as the water's freezing and the ferromagnet's magnetization, unexpected similarity appears not only in the overall pattern of symmetry breaking that these transitions involve, but also in the precise values of parameters—critical exponents—that determine the characteristics of these transitions. When the pressure exerted by water vapor in a container at a fixed temperature increases, the vapor gradually turns into water, going through an intermediate stage at which both gas and fluid are present. In this case, the critical point is a characteristic temperature above which the intermediate stage is no longer manifested and only a homogeneous supercritical fluid is present. The same pattern is observed not only in many other fluids, but also in

ferromagnetic materials, which, above the critical point, lose their magnetization, and ferroelectric materials, which lose the alignment of their electric dipoles. The details of the mechanism clearly differ from case to case; electron spins, for example, play a crucial role in magnetization, but not in freezing or condensation. But the similarity between the systems manifesting universality reveals the overall pattern's insensitivity to structural and dynamic details at the fundamental level.[17] Although it is sometimes questioned whether the theory that explains universality is a physical theory, or merely a mathematical technique,[18] the situation with regard to reduction is quite similar to that of reduction in statistical mechanics. Every system exhibiting universality satisfies the requirement that higher-level patterns supervene on underlying microstructures, but the overall patterns and the parameters that characterize the higher-level patterns are not derived solely from the fundamental laws. Statistical mechanics in general, and the phenomenon of universality in particular, provide clear examples of multiple realizability in physics. Delineating the limits of reducibility, they militate against higher-level eliminativism (causal fundamentalism).

The above examples are not the only interlevel relationships in physics. In chapter 6, I discussed Feynman's explanation of the emergence, from the quantum world, of the classical trajectory satisfying the least action principle. Although there is no fundamental quantum law that parallels the classical principle, integration over the quantum possibilities yields, and explains, the classical trajectory. Yet the explanation does not fit the definition of reduction. For one thing, the two theories in question, quantum mechanics and classical mechanics, seem more radi-

17. This insensitivity is thought to reflect the fact that at (or near) critical points there is a change in the nature of the coupling between components of the system and the range of their relevant interactions. Whereas under normal conditions long-distance coupling and correlations can be ignored, at critical points this idealization is no longer valid and all interactions must be taken into account. Calculation of these overwhelmingly complex processes is made possible by the technique known as the renormalization group, which involves iterative coarse graining of the system, with the result that the system's behavior on every coarse-grained level is analogous to the behavior manifested on the preceding (more fine-grained) level. In the course of this iterative process, the differences between levels within the same system, and the differences between the dynamics of different systems, are washed out.

18. See Morrison (2012) and the references cited there.

cally at variance with one another than the theories in the previous examples. For another, Feynman's explanation, invoking summation over probability amplitudes rather than actual processes, differs considerably from the standard conception of causal explanation. Even the construal of Feynman's explanation as exemplifying supervenience—a much weaker relation than reduction—is shaky. The higher-level classical state is not realized exactly by any particular quantum state, as it is in statistical mechanics, and worse, it is not even unanimously accepted that quantum states represent determinate physical states. It is therefore difficult to consider Feynman's account a reduction of classical to quantum mechanics. At the same time, given Feynman's explanation, the classical principle of least action no longer floats mysteriously above the quantum level, but is securely rooted in it. It combines reductive and emergent features in a manner quite different from that proposed by Bedau.

With these examples in mind, we can now turn to arguments that, seeking to establish full-blown reductionism, purport to demonstrate the inconsistency of higher-level causation. Jaegwon Kim is a good representative of this position. He focuses on Davidson's argument against the reducibility of the mental, but if valid, his arguments should also apply, mutatis mutandis, to the interlevel relations in physics.

Kim sees nonreductivism as dualism, albeit a dualism of properties, not substances:

> Nonreductive physicalism . . . consists of two characteristic theses of non-reductionism: its ontology is physical monism, the thesis that physical entities and their mereological aggregates are all that there is; but its "ideology" is anti-reductionist and dualist, consisting in the claim that psychological properties are irreducibly distinct from the underlying physical and biological properties. Its dualism is reflected in the belief that, though physically irreducible, psychological properties are genuine properties nonetheless, as real as underlying physical-biological properties.
>
> I shall argue that non-reductive physicalism and its more generalized companion, emergentism, are vulnerable to similar difficulties [i.e., similar to those of traditional dualism, YBM]; in particular it

> will be seen that the physical causal closure remains very much a problem within the stratified ontology of non-reductivism. Non-reductive physicalism, like Cartesianism, founders on the rocks of mental causation. (1993, 339)

Furthermore, Kim construes his opponent as claiming that "mentality . . . takes on a causal life of its own and begins to *exercise causal influence 'downward' to affect what goes on in the underlying physical-biological processes*" (1993, 349; emphasis in original). Kim is arguing, then, that the danger ensuing from nonreductive physicalism is downward (higher-level to lower-level) causation. Let us take a closer look at this argument. Suppose, with Kim, that a higher-level property *M* is causally efficacious with respect to another higher-level property *M** and suppose further that these higher-level properties are instantiated by fundamental properties *P* and *P**. According to Kim,

> We seem to have two distinct and independent answers to the question, "Why is this instance of M* present?" *Ex hypothesi*, it is there because an instance of M caused it; that's why it's there. But there is another answer: it's there because P* physically realizes M* and P* is instantiated on this occasion. I believe these two stories about the presence of M* on this occasion create a tension. (Kim 1993, 351)

He continues:

> Is it plausible to suppose that the joint presence of M and P* is responsible for the instantiation of M*? No; because this contradicts the claim that M* is physically realized by P*. . . . This claim implies that P* alone is sufficient to bring about M*. (1993, 352)

Here, Kim conflates the relation of instantiation, which, for any specific case, is an *identity*, with that of causation, which is a relation between two different events. *P** is not, as Kim contends, the cause sufficient to bring about *M**, but rather an instantiation of *M**. The cause of *P** (on Kim's own assumptions) is a *different* microevent, *P*, and the relation between them may be lawful and deterministic. *P* (rather than *P**) is therefore also the cause of *this* instance of *M**. So there is no overdeter-

mination. As we saw, it is possible, and consistent with fundamental physics, that the causal relation between *P* and *P** will not provide a good account of the relation between *M* and *M**. This does not attest to any explanatory lacunae at the basic level, but merely to a change in our explanandum. We have moved from explaining the relation between *P* and *P** to explaining the relation between *M* and *M**. The latter relation, involving different types of events than the former, might be governed by different laws.

Kim, perhaps misled by the upward/downward metaphor, has saddled his opponent with an image of the higher level/basic level structure as akin to a duplex whose upstairs occupants meddle in the affairs of their downstairs counterparts. (Note that here it is the reductionist who slips back into dualism!) This account is a misunderstanding of Davidson; there are no interacting neighbors. Higher levels, as levels of description, are linked to lower levels by various kinds of identities, not by causal connections that can intervene in the causal network at the lower levels.

Kim brings another argument against the causal efficacy of higher-level properties. It is based on a principle he calls the "Causal Inheritance Principle":

> If mental property M is realized in a system at t in virtue of physical realization base P, the causal powers of *this instance of M* are identical with the causal powers of P. (1993, 326; emphasis in original)

In one sense the principle is trivial. The causal powers of this instance of *M* are indeed the causal powers of the physical state that "realizes" it, but this is true simply because this instance of *M is* a *P* state, so that there is only one entity exerting whatever causal influence it has. The idiom of inheritance, though, is misleading, suggesting two distinct entities, one of which inherits something from the other. Might the inheritance principle have a less trivial formulation, for instance, the principle that the causal powers of *M* are "inherited by" every physical state that realizes it? But on this reading, the principle is wrong. In statistical mechanics, we saw, the causal efficacy of macrostates qua macrostates (and their explanatory import) is not inherited by every

microstate that realizes them. My conclusion is that, Kim's arguments notwithstanding, nonreductive physicalism is perfectly consistent with the physical closure of the fundamental level.[19]

Having outlined the limits of reduction in physics, I would like to return to Ryle's claim that "not all questions are physical questions," or to the parallel claim, more germane to the subject of reduction, that not all concepts are physical concepts. Many physicalists take the existence of nonphysical concepts for granted, but since radical physicalists (reductionists) have challenged the existence of such concepts, insisting on their reducibility to physical concepts, the issue deserves our attention.[20] Indeed, the issue has intrigued not only scientists and philosophers but also writers such as Jorge Luis Borges. In several of his stories and essays, Borges explored the question of how the complex world of experience is, or could be, conceptualized.[21] One of these essays is devoted to John Wilkins's attempt to invent an "analytical language."

> He divided the universe into forty categories or classes, which were then subdivided into differences, and subdivided in turn into species. To each class he assigned a monosyllable of two letters; to each difference, a consonant; to each species, a vowel. For example, de means element; deb, the first of the elements, fire; deba, a portion of the element of fire, a flame. . . .

19. Another point to consider is the following. An object's causal efficacy may depend on what happens to *other* objects. E.g., the efficacy of a particular copy of a book may depend on what happens to other copies. If, due to some contingency, only one copy is extant, this copy becomes rare and perhaps more valuable, without any change taking place in the extant copy itself. It might be objected that on this scenario, even if the extant copy has not changed, the microstate of the environment that includes its readers, or the physical world in its entirety, has changed. But if so, we are no longer within the boundaries of a closed system, and the assumption of physical closure is no longer justified.

20. See, e.g., Hemmo and Shenker (2016a); Shenker (2016).

21. To mention a few: "Funes, His Memory," "Averroës' Search," "Pierre Menard, Author of the Quixote." All three first appeared in the 1940s and are translated in Borges's *Collected Fictions* (1998). The essay "John Wilkins' Analytical Language" [1942] is translated in Borges's *Selected Non-Fictions* (1999). Note, however, that with regard to Borges, the division into fiction and nonfiction is artificial; the Wilkins essay, e.g., has features that are patently characteristic of fiction.

> The words of John Wilkins' analytical language are not dumb and arbitrary symbols; every letter is meaningful, as those of the Holy Scriptures were for the Kabbalists. (Borges 1999, 230).

Borges alludes to the fundamental riddle of language: is there a natural way of organizing the world, or is language arbitrary and conventional? We can distinguish between two aspects of the problem: the correspondence between a symbol and what it symbolizes, and the deeper problem of the nature of categories. Borges alludes to the former by comparing Wilkins's analytical language to the Kabbalists' belief in the possibility of a correct, nonconventional language whose terms express the essence of things. With typical irony, Borges describes a lady who declares that "the word *luna* is more (or less) expressive than the word *moon*" (Borges 1999, 229). But even those who fully accept the conventionality of language, in the sense of the arbitrariness of its symbols, must contend with the problem of whether categories are natural or conventional. Borges illustrates the conundrum by adducing a classification of animals, allegedly found in a Chinese encyclopedia—*The Heavenly Emporium of Benevolent Knowledge*:

> In its distant pages it is written that animals are divided into (a) those that belong to the emperor; (b) embalmed ones; (c) those that are trained; (d) suckling pigs; (e) mermaids; (f) fabulous ones; (g) stray dogs; (h) those that are included in this classification; (i) those that tremble as if they were mad; (j) innumerable ones; (k) those drawn with a very fine camel's-hair brush; (l) etcetera; (m) those that have just broken the flower vase; (n) those that at a distance resemble flies. (Borges 1999, 231)

Michel Foucault's *The Order of Things* opens with his reaction to the Chinese encyclopedia's taxonomy:

> This book first arose out of a passage in Borges, out of the laughter that shattered, as I read the passage, all the familiar landmarks of my thought—our thought, the thought that bears the stamp of our age and our geography—breaking up all the ordered surfaces and all the

> planes with which we are accustomed to tame the wild world of existing things, and continuing long afterwards to disturb and threaten with collapse our age-old distinction between the Same and the Other. (Foucault 1970, xv)

Foucault welcomed the destabilizing message of Borges's essay. By telling the history of various conceptions of the word-world relation, Foucault not only presented alternatives to our own conception of the relation between word and object, but also challenged the view that any particular conception is superior to alternative conceptions, let alone that it is correct.

Scientists would also be amused by Borges's fanciful taxonomy, not because they draw the conclusion Foucault drew, but rather, because they seek a "natural classification" that "carves nature at its joints."[22] The problem with Borges's categories is not simply that some of them are empty. There are indeed (or could be) animals belonging to the emperor or animals that have just broken the flower vase. The problem is that such categories are useless from the scientific point of view: not figuring in any laws of nature, they do not underpin predictions and explanations.

The essential correspondence between scientific laws and the concepts (predicates, types, descriptions) they invoke has been repeatedly stressed in this book. Russell's argument to the effect that determinism can be trivially satisfied fails (as we saw in chapter 2) because the putative mathematical function that he takes to exist and to sustain determinism, does not constitute a scientific law. This point has been elaborated on, in a Davidsonian vein, in this chapter. Scientific explanation is sensitive to the descriptions we use and the categories these descriptions refer to. Description sensitivity does not render science subjective or arbitrary. On the contrary, by identifying the categories linked by natural laws, we can distinguish scientific representations of reality from nonscientific representations.

22. The term "natural classification," highlighting the nonarbitrariness of scientific categories, is from Duhem ([1906] 1954, 26). The "carving nature at its joints" metaphor, though not the exact phrase, originates in Plato's *Phaedrus*.

In *Fact, Fiction, and Forecast,* Goodman (1955) pointed to *projectability* as the salient feature of scientific laws, the feature that distinguishes them from accidental regularities. The projectability of laws, however, is inseparable from the entrenchment of the predicates they use. Through his celebrated grue paradox, Goodman shows that "all emeralds are grue," though at first glance analogous to "all emeralds are green," fails the projectability test due to the poor entrenchment of "grue." For simplicity, we can also speak of the projectability of concepts, referring to their appearance in projectable laws.[23]

There is no need to go as far as Goodman's "grue"-some predicates, or to confine the discussion to the mental, as per Davidson's anomalous monism, to find examples of concepts (predicates) that defy lawlikeness and projectability. Consider a stop sign. It is certainly a physical object and belongs to the category of physical objects. It satisfies the laws of physics and does not threaten to violate any laws or confound causal relations at the fundamental level. Nevertheless, there is no physical category that corresponds to the higher-level category of stop signs. This is not just because stop signs are multiply realizable (which, of course, they are), but because the concept of a stop sign is open ended and nonprojectable.[24] Any number of objects could serve as stop signs, and no physical property, structure, or set of specific laws, distinguishes stop signs from other objects. Examples of this kind suggest a refinement of Davidson's point about the mental. It is not a dramatic refinement that is needed, since, after all, stop signs are symbols and require an interpreting mind to understand them. Their open-endedness thus derives from their symbolic significance, and is ultimately predicated on mental activity. Such examples do suggest, however, that the crucial feature differentiating the lawful from the lawless in this context is symbolic meaning rather than mentality per se. Even if mental states such

23. See Abe Stone's insightful "On Scientific Method as a Method for Testing the Legitimacy of Concepts" (Stone 2009). It notes, e.g., that even the paradigmatic example of a black raven may not pass the test for scientific concepts.

24. Multiple realizability in itself does not entail open-endedness. Universality, as we saw, is linked to multiple realizability, but conceivably, is exhibited only in specific kinds of systems. If so, unlike the concept of a stop sign, it is not open ended.

as fear, surprise, and so on were found to correspond to neurological types (which is, perhaps, not unreasonable) or physical types (which is far less plausible), the property of being frightening or surprising would still be open ended and lawless. The same open-endedness and lawlessness also applies to the events and objects falling under these descriptions, for instance, unexpected meetings or frightening movies.

Looking back at the examples discussed in this chapter, we can distinguish several kinds of problems confronting the reductionist assumption of the overarching sovereignty of the fundamental level. First, there were Bedau's examples of deterministic systems that are lawless in the sense that there are no short-cut laws representing their behavior. These cases exhibited neither multiple realizability nor stratification into different conceptual levels. Second, there were examples from physics where higher-level phenomena are subsumed under higher-level concepts and laws foreign to the fundamental level. Here stratification, supervenience, multiple realization, and the insensitivity of higher-level patterns to lower-level detail play an essential role. While these cases do not illustrate lawlessness tout court, they do highlight, on the one hand, the limited explanatory import of fundamental laws (or fundamental causal relations), and on the other, the indispensability of higher-level laws (or higher-level causal relations). This class of cases served to undermine higher-level eliminativism (causal fundamentalism). Lastly, there was a radical mode of lawlessness manifested by open-ended, unprojectable concepts. All these cases were shown to be compatible with determinism and the physical closure assumption. Indeterministic theories like quantum mechanics, however, clearly leave room for further lawlessness.

It is tempting to raise the question of whether these conclusions have bearing on the perennial problem of human freedom. Although, at this time, our best fundamental theory is indeterministic, the macroscopic world with which we interact appears to be governed by classical theories. The idea of a deterministic world in which there is no real freedom therefore continues to haunt us. Determinists often respond to this concern by adducing compatibilism—the view that determinism and freedom are compatible. But compatibilism is based on a *redefinition* of the

notion of freedom. According to the ordinary understanding of freedom, it consists in the existence of genuine alternatives: agents are free when they could have acted otherwise than they in fact did. By contrast, the compatibilist defines free acts as acts that take place in accordance with one's will. Even though (on the assumption of determinism) both the will itself and the acts that accord with it are determined by natural laws, as long as the acts are not carried out against one's will, they are considered to be free. Definitions being (to borrow Dedekind's catchphrase[25]) "free creations of the human mind," compatibilists may be entitled to define freedom as they see fit, but it is doubtful whether this maneuver suffices to solve the problems of moral responsibility, personhood, and so on. Recall Quine on deviant logic. "Here, evidently, is the deviant logician's predicament: when he tries to deny the doctrine, he only changes the subject" (1970, 81). The Davidsonian argument elaborated on in this chapter suggests another option.

Although stop signs are physical objects, the *concept* of a stop sign, I have argued, is not a physical concept. Designing a new type of stop sign (under this description, with this intention in mind, and so on) might likewise not be subject to physical law. This "lawlessness" does not mean that there is no deterministic process leading a particular individual to create the particular sign she is about to create, but that the creation of a sign is not covered by a projectable law that would enable its prediction. The failure of predictability in this case is a matter of principle, entailed by the failure of projectability. Lawlessness of this kind, though compatible with lower-level determinism, makes room for uniqueness and unpredictability in a deterministic world. These qualities are at the heart of our concern about freedom. More than we resent the idea that we are led to act by causes we don't control,[26] we resent the idea that our actions and thoughts are dictated by general laws, are mere instances, that they could have been predicted, that there is nothing unique about them or about us. Protecting uniqueness and unpredictability goes a long way toward satisfying our desire for freedom.

25. Dedekind ([1988] 1996), 791.

26. Recall that we wouldn't want to do away with causation in this context, as we wouldn't want our actions to be random.

Davidson's approach suggests that we can safeguard these aspects of freedom without invoking chance and without engendering any conflict with determinism. Like compatibilism, lawlessness does not give us freedom in the libertarian sense, but it does go beyond compatibilism in making a claim about the world rather than about words, asserting the existence of events ungoverned by law, rather than merely redefining "freedom." In light of these considerations, might not lawlessness be a better option than traditional compatibilism?

REFERENCES

Aharonov, Y., and D. Bohm. 1959. "Significance of Electromagnetic Potentials in the Quantum Theory." *Physical Review* 115: 485–89.

Aharonov, Y., and D. Rohrlich. 2005. *Quantum Paradoxes: Quantum Theory for the Perplexed*. Weinheim: Wiley-VCH.

Albert, D. Z. 2000. *Time and Chance*. Cambridge, MA: Harvard University Press.

Albert, D. Z. 2011. "Physics and Chance." In Ben-Menahem and Hemmo, *Probability in Physics*, 17–40.

Albert, D. Z. 1992. *Quantum Mechanics and Experience*. Cambridge, MA: Harvard University Press.

Alga, K., J. Barnes, J. Mansfeld, and M. Schofield, eds. 1999. *The Cambridge History of Hellenistic Philosophy*. Cambridge: Cambridge University Press.

Anscombe, G.E.M. 1971. *Causality and Determination*. Cambridge: Cambridge University Press.

Aristotle. 1984. *Eudemian Ethics.* In *The Complete Works of Aristotle*, vol. 2, edited by J. Barnes, 1922–1982. Princeton, NJ: Princeton University Press.

Barnum, H., C. M. Caves, J. Finkelstein, C. A. Fuchs, and R. Schack. 2000. "Quantum Probability from Decision Theory?" *Proceedings of the Royal Society of London A* 456: 1175–90.

Batterman, R. W. 2002. *The Devil in the Details*. Oxford: Oxford University Press.

Bedau, M. A. 2008. "Downward Causation and Autonomy in Weak Emergence." In *Emergence*, edited by M. A. Bedau and P. Humphreys, 155–88. Cambridge, MA: MIT Press.

Beebee, H., C. Hitchcock, and P. Menzies, eds. 2009. *The Oxford Handbook of Causation*. Oxford: Oxford University Press.

Bell, J. S. 1976. "How to Teach Special Relativity." *Progress in Scientific Culture* 1 (2). In Bell, *Speakable and Unspeakable in Quantum Mechanics*, 67–80.

Bell, J. S. 1964. "On the Einstein-Podolsky-Rosen Paradox." *Physics* 1 (3): 195–200. In Bell, *Speakable and Unspeakable in Quantum Mechanics*, 14–21.

Bell, J. S. 1966. "On the Problem of Hidden Variables in Quantum Mechanics." *Review of Modern Physics* 38: 447–52. In Bell, *Speakable and Unspeakable in Quantum Mechanics*, 1–13.

Bell, J. S. 2004. *Speakable and Unspeakable in Quantum Mechanics*. 2nd ed. Cambridge: Cambridge University Press.

Belot, G. 1998. "Understanding Electromagnetism." *British Journal for the Philosophy of Science* 49: 531–55.

Ben-Menahem, Y. 2009. "Historical Necessity and Contingency." In *A Companion to the*

Philosophy of History and Historiography, edited by A. Tucker, 120–30. Chichester, UK: Wiley-Blackwell.

Ben-Menahem, Y. 2012. "Locality and Determinism: The Odd Couple." In Ben-Menahem and Hemmo, *Probability in Physics*, 149–65.

Ben-Menahem, Y. 2017. "The PBR Theorem: Whose Side Is It On?" *Studies in the History and Philosophy of Modern Physics* 57: 80–88.

Ben-Menahem, Y., and M. Hemmo, eds. 2012. *Probability in Physics*. Berlin: Springer.

Berlekamp, E. R., J. H. Conway, and R. K. Guy. 1982. *Winning Ways for Your Mathematical Plays*. Vol. 2. New York: Academic Press.

Berlin, I. [1954] 1969. "Historical Inevitability." In *Four Essays on Liberty*, 41–117. London: Oxford University Press.

Birkhoff, G. D. 1931. "Proof of the Ergodic Theorem." *Proceedings of the National Academy of Sciences of the United States of America* 17: 656–60.

Birkhoff, G. D., and J. von Neumann. 1936. "The Logic of Quantum Mechanics." *Annals of Mathematics* 37: 823–45.

Blum, A. 2014. "From the Necessary to the Possible: The Genesis of the Spin-Statistics Theorem." *European Physical Journal* 39: 543–74.

Bobzien, S. 1998. *Determinism and Freedom in Stoic Philosophy*. Oxford: Clarendon.

Bohr, N. [1928] 1983. "The Quantum Postulate and Recent Developments of Atomic Theory." *Nature* 121: 580–90. In *Collected Works*, vol. 6, edited by E. Rudinger. Amsterdam: North Holland. Reprinted in Wheeler and Zurek, *Quantum Theory and Measurement*, 87–126.

Bohr, N. 1935. "Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?" *Physical Review* 48: 696–702.

Borges, J. L. 1998. *Collected Fictions*. Translated by A. Hurley. New York: Penguin.

Borges, J. L. 1999. *Selected Non-Fictions*. Translated by E. Weinberger. New York: Penguin.

Brading, K., and H. R. Brown. 2003. "Symmetries and Noether's Theorems." In Brading and Castellani, *Symmetries in Physics*, 89–109.

Brading, K., and E. Castellani, eds. 2003. *Symmetries in Physics*. Cambridge: Cambridge University Press.

Broadie, S. 2007. *Aristotle and Beyond*. Cambridge: Cambridge University Press.

Bub, J., and I. Pitowsky. 2010. "Two Dogmas about Quantum Mechanics." In *Many Worlds? Everett, Quantum Theory, and Reality*, edited by S. Saunders, J. Barrett, A. Kent, and D. Wallace, 433–59. Oxford: Oxford University Press.

Burbury, S. H. [1894] 1895. "Boltzmann's Minimum Function." Letters to Editor. *Nature* 51: 78, 320.

Butterfield, J. 2011. "Emergence, Reduction, and Supervenience: A Varied Landscape." *Foundations of Physics* 41: 920–59.

Carnot, N.L.S. 1824. *Réflexions sur la puissance motrice du feu et sur les machines propres à développer cette puissance*. Paris: Bachelier.

Carr, E. H. 1961. *What Is History?* London: Macmillan.

Cartwright, N. 1983. *How the Laws of Physics Lie*. Oxford: Clarendon.

Cartwright, N. 2004. "Causation: One Word, Many Things." *Philosophy of Science* 71: 805–19.

Chaitin, G. J. 1966. "On the Length of Programs for Computing Finite Binary Sequences." *Journal of the Association of Computing Machinery* 13: 547–69.

Chaitin, G. J. 1999. *The Unknowable*. Singapore: Springer.

Chandrasekhar, S. 1931a. "The Density of White Dwarf Stars." *Philosophical Magazine*, 7th series, 11: 592–96.

Chandrasekhar, S. 1931b. "The Maximum Mass of Ideal White Dwarfs." *Astrophysical Journal* 74: 81–82.

Chang, H., and N. Cartwright. 1993. "Causality and Realism in the EPR Experiment." *Erkenntnis* 38: 169–90.

Cicero. 1942. *On Fate*. Loeb Classical Library 349. Cambridge, MA: Harvard University Press.

Clatterbaugh, K. 1999. *The Causation Debate in Modern Philosophy, 1637–1739*. London: Routledge.

Conway, J. H., and S. Kochen. 2006. "The Free Will Theorem." *Foundations of Physics* 36: 1441–73.

Cox, B., and J. Forshaw. 2011. *The Quantum Universe: Everything That Can Happen Does Happen*. London: Allen Lane.

Cushing, J. T. 1994. *Quantum Mechanics: Historical Contingency and the Copenhagen Hegemony*. Chicago: University of Chicago Press.

Davidson, D. [1967] 1980. "Causal Relations." In Davidson, *Essays on Actions and Events*, 149–62. Oxford: Clarendon.

Davidson, D. [1970] 1980. "Mental Events." In Davidson, *Essays on Actions and Events*, 207–24. Oxford: Clarendon.

Dedekind, J.W.R. [1988] 1996. "Was sind und was sollen die Zahlen?" In *From Kant to Hilbert*, vol. 2, edited by W. B. Ewald, 790–833. Oxford: Clarendon.

Demopoulos, W. Forthcoming. *Theories*. Cambridge, MA: Harvard University Press.

Demopoulos, W., and I. Pitowsky, eds. 2006. *Physical Theory and Its Interpretation*. Dordrecht: Springer.

Dennett, D. C. 1995. *Darwin's Dangerous Idea*. New York: Simon and Schuster.

Descartes, R. 1985. *The Philosophical Writings of Descartes*. Translated by J. Cottingham, R. Stoothoff, and D. Murdoch. Vols. 1 and 2. Cambridge: Cambridge University Press.

Descartes R. [1641] 1985. *Meditations on First Philosophy*. In Descartes, *Philosophical Writings of Descartes*, vol. 2: 1–62.

Descartes, R. [1644] 1985. *Principles of Philosophy*. In Descartes, *Philosophical Writings of Descartes*, vol. 1: 177–291.

Dirac, P.A.M. 1926. "On the Theory of Quantum Mechanics." *Proceedings of the Royal Society A* 112: 661–77.

Dirac, P.A.M. 1933. "The Lagrangian in Quantum Mechanics." *Physikalische Zeitschrift der Sowjetunion* 3: 64–72. Reprinted in *Feynman's Thesis*, edited by L. M. Brown, 111–19. Singapore: World Scientific, 2005.

DiSalle, R. 1995. "Spacetime Theory as Physical Geometry." *Erkenntnis* 42: 317–37.

Dowe, P. 1992. "Wesley Salmon's Process Theory of Causality and the Conserved Quantity Theory." *Philosophy of Science* 59: 195–216.

Dowe, P. 2000. *Physical Causation*. New York: Cambridge University Press.

Duhem, P. [1906] 1954. *The Aim and Structure of Physical Theory*. Princeton, NJ: Princeton University Press.

Earman, J. 1986. *A Primer on Determinism*. Dordrecht: Reidel.

Earman, J. 2003. "Tracking Down Gauge: An Ode to the Constrained Hamiltonian Formalism." In Brading and Castellani, *Symmetries in Physics*, 140–62.

Earman, J., and J. D. Norton. 1987. "What Price Spacetime Substantivalism? The Hole Story." *British Journal for the Philosophy of Science* 38: 515–25.

Einstein, A. 1905. "Ist die Trägheit eines Körpers von seinem Energieinhalt abhängig?" *Annalen der Physik* 18: 639–41. In *The Collected Papers of Albert Einstein*, vol. 2, doc. no. 24, edited by J. Stachel, 311–15. Princeton, NJ: Princeton University Press. Translated by A. Beck as "Does the Inertia of a Body Depend upon Its Energy Content?" In the English translation of vol. 2: 172–75.

Einstein, A. 1919. "Time, Space, and Gravitation." *The Times* (London), November 28. Reprinted in *The Collected Papers of Albert Einstein*, vol. 7, doc. no. 26, edited by D. Kormos Buchwald et al., 213–15. Princeton, NJ: Princeton University Press.

Einstein, A. 1922. *The Meaning of Relativity*. Princeton, NJ: Princeton University Press. Reprinted in *The Collected Papers of Albert Einstein*, vol. 7, doc. no. 71, edited by D. Kormos Buchwald et al., 496–577. Princeton, NJ: Princeton University Press.

Einstein, A. [1927] 1954. "The Mechanics of Newton and Their Influence on the Development of Theoretical Physics." In *Ideas and Opinions*, 253–61. New York: Crown.

Einstein, A. 1927a. "Bestimmt Schrödingers Wellenmechanik die Bewegung eines Systems vollständig oder nur im Sinne der Statistik?" Unpublished manuscript, forthcoming in *Albert Einstein's Collected Papers*, vol. 15, doc. no. 516, edited by D. Kormos Buchwald et al. Princeton, NJ: Princeton University Press.

Einstein, A. 1931. "About Free Will." In *The Golden Book: Tagore*, edited by R. Chatterjee, 12. Calcutta: Golden Book Committee.

Einstein, A., and M. Grossmann. 1913. "Entwurf einer verallgemeinerten Relativitätstheorie und einer Theorie der Gravitation." Leipzig: B. G. Tübner. In *The Collected Papers of Albert Einstein*, vol. 4, doc. no. 13, edited by M. Klein et al., 302–343. Princeton, NJ: Princeton University Press.

Einstein, A., B. Podolsky, and N. Rosen. 1935. "Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?" *Physical Review* 47: 777–80.

Euler, L. 1744. *Methodus inveniendi lineas curvas maximi minimive proprietate gaudentes* (A Method for Finding Curved Lines Enjoying Properties of Maximum or Minimum). Lausanne: Marcum-Michaelem Bousquet.

Euler, L. [1770] 1833. *Letters to a German Princess*. New York: J. and J. Harper.

Fair, D. 1979. "Causation and the Flow of Energy." *Erkenntnis* 14: 219–50.

Feynman, R. P. 1963. *The Feynman Lectures on Physics*. Vol. 1. San Francisco: Addison-Wesley.

Feynman, R. P. 1965. "The Development of the Space-Time View of Quantum Electrodynamics" (Nobel Lecture). *Physics Today* 19: 31–44.

Feynman, R. P., and A. R. Hibbs. 1965. *Quantum Mechanics and Path Integrals*. New York: McGraw-Hill.

Fodor, J. 1974. "Special Sciences (Or: The Disunity of Science as a Working Hypothesis)." *Synthese* 28: 97–115.

Foucault, M. 1984. "What Is Enlightenment?" In *The Foucault Reader*, edited by P. Rabinow, 32–50. New York: Pantheon.

Frisch, M. 2009a. "Causality and Dispersion: A Reply to John Norton." *British Journal for the Philosophy of Science* 60: 487–95.

Frisch, M. 2009b. " 'The Most Sacred Tenet'? Causal Reasoning in Physics." *British Journal for the Philosophy of Science* 60: 459–74.

Frisch, M. 2014. *Causal Reasoning in Physics*. Cambridge: Cambridge University Press.

Frucht, R. 1939. "Herstellung von Graphen mit vorgegebener abstrakter Gruppe." *Compositio Mathematica* 6: 239–50.

Fuchs, C. 2014. "An Introduction to QBism with an Application to the Locality of Quantum Mechanics." *American Journal of Physics* 82: 749–60.

Gardner, M. 1970. "Mathematical Games—The Fantastic Combinations of John Conway's New Solitaire Game 'Life.' " *Scientific American* 223: 120–23.

Glasner, R. 2009. *Averroes' Physics*. Oxford: Oxford University Press.

Gleason, A. M. 1957. "Measurement on Closed Subspaces of a Hilbert Space." *Journal of Mathematics and Mechanics* 6: 885–93.

Glennan, S. 2009. "Mechanisms." In Beebee, Hitchcock, and Menzies, *Oxford Handbook of Causation*, 315–25.

Godfrey-Smith, P. 2009. "Causal Pluralism." In Beebee, Hitchcock, and Menzies, *Oxford Handbook of Causation*, 326–37.

Goldstein, H. [1950] 1973. *Classical Mechanics*. Reading, MA: Addison-Wesley.

Goldstein, S., D. V. Tausk, R. Tumulka, and N. Zanghì. 2009. "What Does the Free Will Theorem Actually Prove?" arXiv: 0905.4641v1 [quant-ph].

Goodman, N. 1955. *Fact, Fiction, and Forecast*. Cambridge, MA: Harvard University Press.

Gould, S. J. 1989. *Wonderful Life*. New York: W. W. Norton.

Haag, R. 1993. *Local Quantum Physics*. Berlin: Springer.

Harris, C. L. 1981. *Evolution, Genesis, and Revelations: With Readings from Empedocles to Wilson*. Albany: State University of New York Press.

Hart, H.L.A., and A. M. Honoré. 1959. *Causation in the Law*. Oxford: Oxford University Press.

Healey, R. 2007. *Gauging What's Real*. Oxford: Oxford University Press.

Hegel, G.W.F. [1835–38] 1975. *Lectures on the Philosophy of World History*. Translated from the German by H. S. Nisbet. Cambridge: Cambridge University Press.

Heisenberg, W. 1926. "Mehrkörperprobleme und Resonanz in der Quantenmechanik." *Zeitschrift für Physik* 38: 411–26.

Heisenberg, W. 1930. *The Physical Principles of Quantum Theory*. Chicago: University of Chicago Press.

Hemmo, M., and O. Shenker. 2012a. "The Mathematical Representation of the Arrow of Time." *Iyyun* 61: 167–92.

Hemmo, M., and O. Shenker. 2012b. "Measure over Initial Conditions." In Ben-Menahem and Hemmo, *Probability in Physics*, 87–98.

Hemmo, M., and O. Shenker. 2012c. *The Road to Maxwell's Demon*. Cambridge: Cambridge University Press.

Hemmo, M., and O. Shenker. 2016. "A Physicalist Account of Multiple Realizability in the Special Sciences." philsci-archive.pitt.edu/12619.

Hitchcock, C. 2007. "What Russell Got Right." In Price and Corry, *Causation, Physics, and the Constitution of Reality*, 45–65.

Hume, David. [1739] 1978. *A Treatise of Human Nature*. Edited by L. A. Selby-Bigge and P. H. Niditch. Oxford: Clarendon.

Hume, David. [1748] 2000. *An Enquiry concerning Human Understanding*. Edited by T. L. Beauchamp. Oxford: Clarendon.

James, W. [1884] 1956. "The Dilemma of Determinism." In *The Will to Believe*, 145–83. New York: Dover.

Jaynes, E. T. 1983. *Papers on Probability, Statistics, and Statistical Physics*. Dordrecht: Reidel.

Kant, I. 1793. *Die Religion innerhalb der Grenzen der bloßen Vernunft*. Königsberg: Friedrich Nicolovius.

Kim, J. 1993. *Supervenience and Mind*. Cambridge: Cambridge University Press.

Kochen, S., and E. P. Specker. 1967. "The Problem of Hidden Variables in Quantum Mechanics." *Journal of Mathematics and Mechanics* 17: 59–89.

Kreisel, G. 1958. "Wittgenstein's Remarks on the Foundation of Mathematics." *British Journal for the Philosophy of Science* 9: 135–58.

Kvart, I. 1986. *A Theory of Counterfactuals*. Indianapolis: Hackett.

Lagrange, J. L. [1811] 1997. *Analytical Mechanics*. Translated by A. Boissonnade and V. N. Vagliente from the 1811 edition of *Méchanique analitique*. Boston Studies in the Philosophy of Science 191. Dordrecht: Kluwer.

Lanczos, C. 1949. *The Variational Principles of Mechanics*. Toronto: University of Toronto Press.

Lange, M. 2017. *Because without Cause: Non-Causal Explanations in Science and Mathematics*. Oxford: Oxford University Press.

Laplace, P. S. [1814] 1994. *A Philosophical Essay on Probabilities*. Translation of fifth edition by A. I. Dale. New York: Springer.

Leibniz, G. W. [1675–1716] 1989. *Philosophical Essays*. Edited and translated by R. Ariew and D. Garber. Indianapolis: Hackett.

Leibniz, G. W. [1915–16] 1956. *The Leibniz Clarke Correspondence*. Edited by H. G. Alexander. Manchester: University of Manchester Press.

Lewis, D. 1973. "Causation." *Journal of Philosophy* 70: 556–67.

Lo, Hoi-Kwong, T. P. Spiller, and S. Popescu, eds. 1998. *Introduction to Quantum Computation and Information*. Singapore: World Scientific.

Mackie, J. 1965. "Causes and Conditions." *American Philosophical Quarterly* 2: 245–55.

Mahoney, M. S. 1973. *The Mathematical Career of Pierre de Fermat*. Princeton, NJ: Princeton University Press.

Malament, D. 2008. "Norton's Slippery Slope." *Philosophy of Science* 75: 799–816.

Martin, C. A. 2003. "On Continuous Symmetries and the Foundations of Modern Physics." In Brading and Castellani, *Symmetries in Physics*, 29–60.

Massimi, M. 2005. *Pauli's Exclusion Principle: The Origin and Validation of a Scientific Principle*. Cambridge: Cambridge University Press.

Maudlin, T. 1994. *Quantum Non-Locality and Relativity*. Oxford: Blackwell.

Maudlin, T. 2007. *The Metaphysics within Physics*. Oxford: Oxford University Press.

Maupertuis. 1751. *Essai de Cosmologie*. Amsterdam: s.n.

Maxwell, J. C. [1877] 1920. *Matter and Motion*. New York: Dover.

Maynard Smith, J. 1992. "Taking a Chance on Evolution." *New York Review of Books* 39 (May 14): 234–36.

Menzies, P., and H. Price. 1993. "Causation as a Secondary Quality." *British Journal for the Philosophy of Science* 44: 187–203.

Mermin, N. D. 1998. "Nonlocal Character of Quantum Theory?" *American Journal of Physics* 66: 920–24.

Meyerson, E. [1908] 1930. *Identity and Reality*. London: George Allen and Unwin.

Morrison, M. 2003. "Spontaneous Symmetry Breaking: Theoretical Arguments and Philosophical Problems." In Brading and Castellani, *Symmetries in Physics*, 347–64.

Morrison, M. 2012. "Beyond Calculation: Extracting Physical Information from Mathematical Methods." *Iyyun* 61: 149–66.

Myrvold, W. C., and J. Christian, eds. 2009. *Quantum Reality, Relativistic Causality, and Closing the Epistemic Circle*. Dordrecht: Springer.

Nagel, E. 1961. *The Structure of Science*. London: Routledge and Kegan Paul.

Nickles, T. 1973. "Two Concepts of Intertheoretic Reduction." *Journal of Philosophy* 70: 181–201.

Nietzsche, F. [1876] 1980. *On the Advantage and Disadvantage of History for Life*. Translated from the German by P. Preuss. Indianapolis: Hackett.

Norton, J. D. 2007. "Causation as Folk Science." In Price and Corry, *Causation, Physics, and the Constitution of Reality*, 11–44.

Norton, J. D. 2009. "Is There an Independent Principle of Causality in Physics?" *British Journal for the Philosophy of Science* 60: 475–86.

Norton, J. D. 2015. "The Hole Argument." *Stanford Encyclopedia of Philosophy*, fall 2015 edition, edited by Edward N. Zalta. plato.stanford.edu/archives/fall2015/entries/spacetime-holearg/.

Norton, J. D. 2016. "Curie's Truism." *Philosophy of Science* 83: 1014–26.

Nöther, E. 1918. "Invariante Variationsprobleme." *Königliche Gesellschaft der Wissenschaften zu Göttingen. Mathematisch-Physikalische Klasse. Nachrichten*, 235–57. English translation by M. A. Tavel. 1971. "Noether's Theorem." *Transport Theory and Statistical Physics* 1: 183–207.

Oppenheim, J., and S. Wehner. 2010. "The Uncertainty Principle Determines the Nonlocality of Quantum Mechanics." *Science* 330: 1072–74.

Pauli, W. 1925. "Über den Zusammenhang des Abschlusses der Elektronengruppen im Atom mit der Komplexstruktur der Spektren." *Zeitschrift für Physik* 31: 765–83.

Pauli, W. 1940. "The Connection between Spin and Statistics." *Physical Review* 58: 716–22.

Pearl, J. 2000. *Causality*. Cambridge: Cambridge University Press.

Peirce, C. S. 1878. "How to Make Our Ideas Clear." *Popular Science Monthly* 12: 286–302.

Penrose, O., and I. C. Percival. 1962. "The Direction of Time." *Proceedings of the Physical Society* 79: 605–16.

Pitowsky, I. 1989. *Quantum Probability, Quantum Logic, Lecture Notes in Physics 321*. Heidelberg: Springer.

Pitowsky, I. 1996. "Laplace's Demon Consults an Oracle: The Computational Complexity of Prediction." *Studies in History and Philosophy of Modern Physics* 27: 161–80.

Pitowsky, I. 2006. "Quantum Mechanics as a Theory of Probability." In Demopoulos and Pitowsky, *Physical Theory and its Interpretation*, 213–39.

Pitowsky, I. 2012. "Typicality and the Role of the Lebesgue Measure in Statistical Mechanics." In Ben-Menahem and Hemmo, *Probability in Physics*, 41–58.

Planck, M. 1949. *Scientific Autobiography and Other Papers*. New York: Philosophical Library.

Poincaré, H. 1905. *Leçons de mécanique céleste*. Vol. 1. Paris: Gauthier-Villars.

Popescu, S., and D. Rohrlich. 1994. "Quantum Nonlocality as an Axiom." *Foundations of Physics* 24: 379–85.

Popescu, S., and D. Rohrlich. 1998. "The Joy of Entanglement." In Lo, Spiller, and Popescu, *Introduction to Quantum Computation and Information*, 29–48.

Price, H. 1996. *Time's Arrow and Archimedes' Point*. Oxford: Oxford University Press.

Price, H., and R. Corry, eds. 2007. *Causation, Physics, and the Constitution of Reality*. Oxford: Oxford University Press.

Pruss, A. 2006. *The Principle of Sufficient Reason.* Cambridge: Cambridge University Press.

Pusey, M. F., J. Barrett, and T. Rudolph. 2012. "On the Reality of the Quantum State." *Nature Physics* 8: 475–78.

Putnam, H. [1967] 1975a. "The Nature of Mental States." In *Mind, Language, and Reality*, Philosophical Papers, vol. 2: 429–440**.** Cambridge: Cambridge University Press.

Putnam, H. [1973] 1975b. "Philosophy and Our Mental Life." In *Mind, Language, and Reality*, Philosophical Papers, vol. 2: 291–303.

Putnam, H. 1994. "Why Functionalism Didn't Work." In *Words and Life*, edited by J. Conant, 441–59. Cambridge MA: Harvard University Press.

Quine, W.V.O. 1960. *Word and Object*. Cambridge, MA: MIT Press.

Quine, W.V.O. 1966. *The Ways of Paradox and Other Essays*. New York: Random House.

Quine, W.V.O. 1970. *Philosophy of Logic*. Englewood Cliffs, NJ: Prentice-Hall.

Railton, P. 1978. "A Deductive-Nomological Model of Probabilistic Explanation." *Philosophy of Science* 45: 206–26.

Railton, P. 1981. "Probability, Explanation, and Information." *Synthese* 48: 233–56.

Redhead, M. 1987. *Incompleteness, Nonlocality, and Realism*. Oxford: Clarendon.

Redhead, M. 2003. "The Interpretation of Gauge Symmetry." In Brading and Castellani, *Symmetries in Physics*, 124–39.

Reichenbach, H. 1956. *The Direction of Time*. Los Angeles: University of California Press.

Rickles, D. 2008. *Symmetry, Structure, and Spacetime*. Amsterdam: Elsevier.

Rorty, R. 1989. *Contingency, Irony, and Solidarity*. Cambridge: Cambridge University Press.

Rosen, J. 2008. *Symmetry Rules: How Science and Nature are Founded on Symmetry*. Berlin: Springer.

Russell, B. 1913. "On the Notion of Cause." *Proceedings of the Aristotelian Society* 13: 1–26.

Russell, B. 1919. *Introduction to Mathematical Philosophy*. London: George Allen and Unwin.

Russell, B. 1948. *Human Knowledge: Its Scope and Limits*. London: Allen and Unwin.

Ryle, G. 1949. *The Concept of Mind*. London: Hutchinson.

Sabra, A. I. 1967. *Theories of Light from Descartes to Newton*. Cambridge: Cambridge University Press.

Salmon, W. 1984. *Scientific Explanation and the Causal Structure of the World*. Princeton, NJ: Princeton University Press.

Salmon, W. 1989. *Four Decades of Scientific Explanation*. Pittsburgh: University of Pittsburgh Press.

Schlick, M. 1920. Albert Einstein Archive 21-576. *The Collected Papers of Albert Einstein*, vol. 10, doc. no. 51. Princeton, NJ: Princeton University Press.

Schrödinger, E. 1935. "The Present Situation in Quantum Mechanics." Translated by J. D. Trimmer. In Wheeler and Zurek, *Quantum Theory and Measurement*, 152–67.

Schrödinger, E 1952. "Are There Quantum Jumps?" *British Journal for the Philosophy of Science* 3, part 1, 109–23; part 2, 233–42. In *Erwin Schrödinger: Collected Papers*, vol. 4, edited by A. Dick, G. Kerber, and W. Kerber, 478–502. Vienna: Verlag der Oesterreichischen Akademie der Wissenschaften, 1984.

Schweber, S. S. 1994. *QED and the Men Who Made It*. Princeton, NJ: Princeton University Press.

Shenker, O. 2016. "A Reductive Physicalist Account of the Autonomy of Psychology." http://philsci-archive.pitt.edu/12620/.

Shimony, A., and H. Stein. 2003. "On Quantum Non-locality, Special Relativity, and Counterfactual Reasoning." In *Revisiting the Foundations of Relativistic Physics*, edited by A. Ashtekar et al., 499–521. Dordrecht: Kluwer.

Sklar, L. 1993. *Physics and Chance*. Cambridge: Cambridge University Press.

Skyrms, B. 1984. "EPR: Lessons for Metaphysics." *Midwest Studies in Philosophy* 9: 245–55.

Spekkens, R.W. 2005. "In Defense of the Epistemic View of Quantum States: A Toy Theory." arXiv:quant-ph/0401052v2.

Spinoza, B. [1677] 1985. *Ethics*. In *The Collected Works of Spinoza*, vol. 1, edited by E. Curley, 408–620. Princeton, NJ: Princeton University Press.

Stachel, J. 1969. "Comments on 'Causality Requirements and the Theory of Relativity.' " *Boston Studies in the Philosophy of Science* 5: 179–98.

Stachel, J. [1980] 1989. "Einstein's Search for General Covariance, 1912–1915." In *Einstein and the History of General Relativity*, edited by D. Howard and J. Stachel, 63–100. Einstein Studies 1. Boston: Birkhäuser. Slightly revised in Stachel, *Einstein from "B" to "Z,"* 309–25.

Stachel, J. 2002. *Einstein from "B" to "Z."* Boston: Birkhäuser.

Steiner, M. 1986. "Events and Causality." *Journal of Philosophy* 83: 249–64.

Steiner, M. 1998. *The Applicability of Mathematics as a Philosophical Problem*. Cambridge, MA: Harvard University Press.

Stone, A. D. 2009. "On Scientific Method as a Method for Testing the Legitimacy of Concepts." *Review of Contemporary Philosophy* 8: 13–48.

Strickland, L. 2006. *The Shorter Leibniz Texts*. London: Continuum.

Suppes, Patrick. 1970. *A Probabilistic Theory of Causality*. Amsterdam: North-Holland.

Taylor, R. 1974. *Metaphysics*. 2nd ed. Englewood Cliffs, NJ: Prentice-Hall.

Terrall, M. 2002. *The Man Who Flattened the Earth*. Chicago: University of Chicago Press.

Tsirelson (Cirel'son), B. S. 1980. "Quantum Generalizations of Bell's Inequality." *Letters in Mathematical Physics* 4: 93–100.

Tumulka, R. 2006. "A Relativistic Version of the Ghirardi-Rimini-Weber Model." *Journal of Statistical Physics* 125: 821–40.

Uffink, J. 2007. "Compendium to the Foundations of Classical Statistical Physics." In J. Butterfield and J. Earman, eds., *Philosophy of Physics*, 924–1074. Amsterdam: Elsevier.

Vaidman, L. 2012. "On the Role of Potentials in the Aharonov-Bohm Effect." *Physical Review A* 86 (4): 040101.

Van Fraassen, B. C. 1980. *The Scientific Image*. Oxford: Clarendon.

Von Neumann, J. 1932. "Proof of the Quasi-Ergodic Hypothesis." *Proceedings of the National Academy of Sciences of the United States of America* 18: 70–82.

Wallace, D. 2003. "Time Dependent Symmetries: The Link between Gauge Symmetries and Indeterminism." In Brading and Castellani, *Symmetries in Physics*, 163–73.

Weyl, H. [1918] 1952. *Space-Time-Matter*. Translated by Henry L. Brose. New York: Dover.

Weyl, H. 1952. *Symmetry*. Princeton, NJ: Princeton University Press.

Wheeler, J. A., and W. H. Zurek. 1983. *Quantum Theory and Measurement*. Princeton, NJ: Princeton University Press.

White, H. 1973. *Metahistory*. Baltimore: Johns Hopkins University Press.

Wigner, E. P. 1967. *Symmetries and Reflections*. Bloomington: Indiana University Press.

Wolfram, S. 1994. *Cellular Automata and Complexity*. Reading, MA: Addison-Wesley.

Woodward, J. 2003. *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.

Wu, T. T., and C. N. Yang. 1975. "Concept of Nonintegrable Phase Factors and Global Formulation of Gauge Fields." *Physical Review D* 12: 3845–57.

INDEX

A NOTE ON THE TYPE

This book has been composed in Arno, an Old-style serif typeface in the classic Venetian tradition, designed by Robert Slimbach at Adobe.